RATIONALITY & COMPETITION

CRC TRR 190

Does Relative Performance Information Lower Group Morale?

Lea Heursen (HU Berlin)

Discussion Paper No. 213

December 12, 2019

Does Relative Performance Information
Lower Group Morale?

Lea Heursen[*]

This version: December 2019

Abstract

In many organizations, productivity relies not just on individual effort but also on group morale, that is, the willingness of co-workers to help each other perform better at work. Relative performance evaluations (RPE) are known to increase individual work morale but may negatively affect group morale because they create a sense of competition among members of a reference group. In a novel experiment, I vary whether or not members of a reference group obtain relative performance information on a task that is relevant for their social image or self-image, a general knowledge test. I measure how this affects the subsequent willingness to help the productivity of others by sharing knowledge with them at a personal cost. I find that RPE cause members of a reference group to compete as intensely as under relative pay, compared to a baseline with no relative performance information and fixed piece-rates. It also increases the perceived social distance between them. Yet, I show that even after a performance competition, individuals are willing to help the productivity of others in the group. These findings advance our understanding of how relative concerns among co-workers affect the way they work together.

Keywords: relative performance information, rank feedback, social incentives, on-the-job help, group productivity, social and self-image, experiment
JEL codes: D23, C92, J24, D91

1. Introduction

Jobs require more and more that employees work together. (Lazear and Shaw 2007; Deming 2017). Therefore, to understand productivity in organizations, it is important to understand what makes groups of co-workers productive. In many organizations with group work, social comparison in performance is encouraged. Firms such as, for example, Amazon[2] or Yahoo[3], evaluate their employees' relative performances by ranking them. Schools and universities are another class of organizations in which relative performance evaluations are common. These type of social incentives are known to increase the productivity of individuals (Blanes i Vidal and Nossol 2011; Kuhnen and Tymula 2012; Tran and Zeckhauser 2012; Gill et al. 2018). But little attention has been paid to how they affect other behavior in workgroups. Relative performance information seems to encourage cheating and sabotage in reference groups (Charness, Masclet and Villeval 2014). How do relative performance evaluations affect how well people work together?

One important factor of how well people work together is *group morale* that I define as *the willingness of co-workers to help each other perform better at work.*[4]

Microsoft is an example of a firm where relative performance information appears to have had unintended consequences in workgroups. In November 2013, the company abolished a relative performance evaluation scheme under which employees within organizational units were ranked according to their performances. This ranking system was partly blamed for a "lost decade" at the company[5] and highly unpopular with employees. "It leads to employees focusing on competing with each other rather than competing with other companies", as one Microsoft employee describes it. An important rationale for the abolition was to "promote new levels of teamwork" and "put more emphasis on teamwork and collaboration" (Microsoft Humam Ressource Chief).[6]

In this paper, I use a laboratory experiment to investigate empirically whether this concern is warranted. Does relative performance information lower group morale? The study is designed to shed light on the following mechanism: relative performance information may cause members

---

[2] "Inside Amazon: Wrestling Big Ideas in a Bruising Workplace", The New York Times, accessed April 9, 2018 https://www.nytimes.com/2015/08/16/technology/inside-amazon-wrestling-big-ideas-in-a-bruising-workplace.html

[3] Marissa Mayer famously introduced a relative performance ranking system when she joined Yahoo in 2012, see, e.g., "What Marissa Mayer Got Wrong (and Right) About Stack Ranking Employees", Harvard Business Review, accessed April 9, 2018 https://hbr.org/2015/01/what-marissa-mayer-got-wrong-and-right-about-stack-ranking-employees

[4] I use the term group morale in a narrow sense, to describe the willingness of individuals to help the performance of others. With this, I follow how economists have integrated the concept of individual (work) morale in their work. This concept also emphasizes the motivation to exert more effort in order to perform better individually.

[5] "Microsoft's Lost Decade", Vanity Fair, accessed 9 April, 2018, https://www.vanityfair.com/news/business/2012/08/microsoft-lost-mojo-steve-ballmer

[6] "Microsoft Axes its Controversial Employee-ranking System", The Verge, accessed on 9. April 2018, https://www.theverge.com/2013/11/12/5094864/microsoft-kills-stack-ranking-internal-structure

of a reference group to compete against one another and this may lower their willingness to help each other perform better.

The experimental laboratory offers two important advantages over observational data sets. Firstly, I can isolate the effect that a competitive environment in and of itself has on the willingness to help the productivity co-workers. Career concerns can offer a strategic motive to not invest in the productivity of those who compete for promotions. They are hard to disentangle from relative concerns *per se,* under which co-workers simply want to maintain outcome differences. In the experiment, strategic motives are completely shut down because helping others' productivity does not affect the probability of obtaining a high or low performance rank. Secondly, I can pin down the effect of relative performance information because I can precisely control a major confounding factor present in data from companies such as Microsoft: high relative performances come, at least in the medium run, with a monetary prize such as promotions. Employees may simply compete for the money that comes with high rankings. Alternatively, the relative performance information *per se* may be sufficient to spur a competition for non-monetary or image-based rewards. Given that relative performance information is ubiquitous, it is important to know whether this feature of an organization alone can cause a competition and lead to unintended consequences within reference groups.

In my experiment, a total of 282 subjects participate in one of four experimental conditions. I vary experimentally whether members of a reference group receive only absolute feedback on a timed general knowledge test (baseline) or also an evaluation of their relative performances on this test. In the private rank feedback treatment, a group member observes her performance rank in the group (rank 1, rank 2 or rank 3). In the public rank feedback treatment, I establish common knowledge about performance ranks in the group, by displaying the picture of a group member next to that group member's performance rank. Thus, these two treatments give people the kind of relative performance feedback that has been shown to motivate people to work harder individually (see e.g. Blanes i Vidal and Nossol 2011; Tran and Zeckhauser 2012). A control treatment introduces relative pay. A group member observes her performance rank in private and the best performer in a group earns a substantial monetary bonus.

I use these conditions to test, in a between-subject design, whether relative performance evaluations cause the perception of competition within a reference group and whether this impacts on group members' willingness to help others in their group perform better. With the relative pay control treatment, I explore whether any results change moving from purely image-based rewards to a monetary reward for relative performance.

For this purpose, in the second part of the experiment, I measure group morale in a way that closely resembles the type of help that is important in workgroups: sharing knowledge.

Participants answer more general knowledge questions for a fixed monetary piece-rate. They can invest in the performance of others in their reference group by sharing their answers

to this new set of questions at a small personal cost. Sharing answers can improve the performance and earnings of other group members because the computer automatically replaces their incorrect answers with a correct answer that was shared.

I model helping as a prosocial act. It comes at a small cost and benefits others in the group without any direct *monetary* benefit for the person who lends helps. This mirrors the actual cost of helping in organizations. Moreover, this type of prosocial help among co-workers should be very sensitive to the intensity of competition between them.

Throughout the experiment, nobody observes the help behavior of others. Importantly, this design feature rules out that the desire to demonstrate to others a (perceived) advantage in knowledge or a prosocial attitude motivate knowledge-sharing decisions.

This study introduces the knowledge-sharing task in order to elicit "real help"[7] behavior: the outcome of the help decision depends on the performance of the person who lends help and the one who receives help. The knowledge-sharing task captures this important dimension of on-the-job help outside of the laboratory. This sets it apart from experimental paradigms that model prosocial workplace behavior with the transfer of money.

In the third part of the experiment, I measure participants' beliefs about the correctness of their own and their group members' answers to all questions for which help decisions were made, in an incentivized and incentive-compatible way. This data is important for two reasons. Firstly, I can confirm that people share their knowledge in order to make others perform better. Secondly, I can account for the fact that the rank information of the treatments may change beliefs about the value of own help to others in the group. This may contribute to people helping differently across experimental conditions, in addition to the competition channel. I also elicit the sense of competition and social distance in reference groups, as well as beliefs about the expected help by others. With this data, I evaluate the impact that relative performance evaluations have on the perception of social relations in reference groups.

I find that both private and public relative performance feedback causes a large and statistically significant increase in the level of competition among group members, compared to the baseline condition with a very low level of competition. I find that the intensity of competition under private or public relative performance evaluations, is, on average, as large as the intensity of competition when group members compete for relative pay. Relative performance evaluations also increase the social distance between members of experimental reference groups.

I then show that even after a performance competition, a substantial share of participants helps the productivity of others in the group. With my data, I can very precisely estimate that after a performance competition, group members are, on average, as willing to help each other

---

[7] Analogously to how experimental economists use the term "real effort" (see e.g. Charness, Gneezy, and Henderson 2018) the term "real help" is used to indicate that the outcome of the help decision depends on the performances of the person who lends help and of the person who receives help.

perform better as in the baseline condition in which group members do not compete.[8] This is true for the private rank feedback, the public rank feedback and the relative pay treatments.

Participants are much more likely to share answers that they think will improve the performance of their group members, as evidenced by estimates of binary choice models that predict the probability of sharing answers. This is consistent with a measure of group morale.

My study offers two insights for relative performance evaluations and productivity in organizations. Firstly, relative performance information is likely to trigger a comparative and competitive mind-set in reference groups. Secondly, this study's findings provide initial evidence that the willingness of co-workers to extend on-the-job help does not systematically decrease as the work environment becomes more competitive.

This is a more positive outlook on relative concerns among co-workers and how they impact on the way co-workers interact and work together, than what the prior views of business insiders[9] or related research may suggest.

A small but growing empirical literature finds that relative pay concerns among co-workers entail behavioral spillovers to workplace behavior. In a laboratory experiment with a real effort task, Carpenter, Matthews and Schirm (2010) find that relative performance pay causes an increase in sabotage among members of a reference group, even if sabotaging others' work has no effect on the likelihood of winning the bonus competition. In a large-scale online experiment, Buser and Dreber (2016) show that competing for relative pay significantly and sizably lowers subsequent contributions in an unrelated Public Goods Game, compared to a baseline condition with a fixed piece-rate. Breza, Kaur, and Shamdasani (2018) investigate in a field experiment how relative pay concerns affect workers' individual work morale and their ability to work together. They find that a shared history of wage disparities, established on the basis of relative performance differences in reference groups, lowers co-workers' ability to produce together in joint production tasks in which it is in the self-interest of co-workers to work together. This present study adds to this literature by isolating how relative concerns in a non-monetary domain, the performance domain, affect how well members of a reference group work together. In light of the prominence of relative performance information in organizations, this is an important gap to fill.

---

[8] Given the study's sample size, I can calculate the power of a two-sided t-test at a level of significance of at least $\alpha = 0.05$ to reject the null hypothesis that group morale is not affected by a performance competition in reference groups for different standardized effect sizes (d), i.e. standardized difference in means across two conditions. For what is typically considered a medium standardized effect size $d = 0.5$, this study would reject the null hypothesis of no effect 85% of the times with the likelihood of a Type-2 error as low as 15%.

[9] In addition to the perception at Microsoft, CEOs of other companies have also voiced concerns about relative evaluations of employees on the ground that they may hinder collaboration. Qualtric's CEO Smith is quoted with the following opinion on relative performance evaluations: "Stack-ranking is fine, says Smith, for evaluating performance in a sales organization, where managers may want to heighten competition. It's less well suited, he says, for evaluating engineers, among whom management may want to create closer collaboration.", "Cooperation "Microsoft: 'Stack-Ranking' Gets Heave-Ho", accessed 9, April , 2018 https://abcnews.go.com/Business/microsoft-abolishes-stack-ranking-employees/story?id=20877556

Two treatment manipulations—private and public performance rank feedback on a general knowledge test—build on a large conceptual literature in behavioral economics on people being motivated by self-image (Köszegi 2006) and social image (Bénabou and Tirole 2006; Ellingsen and Johannesson 2008; Besley and Ghatak 2008). Several empirical studies confirm that people like to signal to themselves or to others that they are intelligent (Tran and Zeckhauser 2012, Ewers and Zimmermann 2015). But there are a few studies that find that introducing an audience can actually lower the desire to signal competence or ambition to others (McManus and Rao 2015; Bursztyn, Fujiwara, and Pallais 2017). I considered these findings in the design of my experiment, attempting to ensure that performing well and being seen as performing well at the real effort task is desirable for the participants.

This paper explores a potential cost to relative performance information. Two related studies reveal different unintended side effects of relative performance feedback in firms. In a field experiment, Bandiera, Barankay and Rasul (2013) show that rank feedback reduces the average productivity in a firm because this information changes how work teams form. Charness, Masclet and Villeval (2014) show that public performance rankings cause members of a reference group to spend money in order to artificially increase their own performance, i.e. to cheat, or to lower the output of others in their group, i.e. to sabotage. The findings of my study draw attention to the fact that more research is needed to understand when social incentives in the form of relative performance feedback do, and when they do not backfire in reference groups.

My paper also contributes to the empirical literature on on-the-job help. While a link between tournaments in firms and incentives to help has been established theoretically (Lazear 1989), there are only a few papers in economics that investigate empirically the determinants of on on-the-job help. With survey data among employees of a plant, Drago and Garvey (1998) provide correlational evidence that promotion tournaments lower employees' qualitative ratings of their co-workers willingness to provide help (e.g. by sharing tools or machinery). Danilov, Harbring, and Irlenbusch (2019) present findings from a laboratory experiment that models help as the transfer of money and compares help under different combinations of team-pay and relative performance pay. They find that helping changes with the monetary incentives in the expected direction; it decreases as the relative performance incentive raises and increases as the team-pay gets larger. The findings oft this present study advance this literature by providing causal evidence on whether relative concerns among co-workers in and of themselves systematically affect the willingness to help the productivity of other co-workers.

The rest of the paper is organized as follows, section 2 presents a framework to illustrate the effects under study, in section 3, I outline the design of my experiment in detail. Section 4 shows the results and section 5 discusses them and concludes.

2. Framework

In the following, I present a simple conceptual framework to illustrate how relative performance evaluations can change group morale, that is, the willingness to help others perform better in reference groups. An employee (i) receives a piece-rate b for good performance on a task. The employee either knows how to solve this task ($p_i = 1$) or does not ($p_i = 0$). Her utility is $u_i = bp_i$.

She can lend help $h_i \in \{0,1\}$ to a co-worker (j) in her reference group, for example, by sharing her knowledge on how to solve the task, to improve his performance. Helping comes at a small cost $c > 0$.[10] Her payoff $v(h_i)$ from helping is:

$$v(h_i) = G(a_i, p_i, p_j)u_j(h_i) - ch_i$$

where $G()$ is a group morale effect term. Helping a colleague can increase his performance. I use the notation $u_j(h_i)$ to indicate j's expected monetary payoff as a function of employee i's help. Likewise, $u_i(h_j)$ denotes i's expected monetary payoff as a function of the help of colleague j.[11]

Clearly, the employee helps her co-worker whenever

$$\Delta v(h_i) = G(a_i, p_i, p_j)\Delta u_j(h_i) \geq c.$$

The benefit to help is increasing in $G()$ such that, *ceteris paribus*, employees with a higher group morale are more likely to help other colleagues to perform better. Helping is an act that benefits her co-worker, therefore, an employee's group morale is linked to her prosociality. I conceptualize it in the following way:

$$G(a_i, p_i, p_{-i}) = a_i - s(p_i, p_j)$$

The employee gives weight $a_i \in [0,1]$ to the utility of her co-worker. There is ample evidence that people differ in their prosocial inclinations. Thus, this simply captures that some colleagues are more inclined to help than others and some may never help (e.g. $a_i = 0$). In addition, relative performance concerns $s(p_i, p_j)$ may change an employee's group morale or, in other words, the value that she assigns to improving the performance and earnings of her colleague. Note that without this second term, the group morale term $G()$ is a simple altruism model. It is common to conceptualize the weight that people give to the utility of others as partly determined by the decision context, for example, by others' behavior (Rabin 1993) or by their prosocial type (Levine 1998) or, as in this framework, by relative performance concerns.

---

[10] For example, there is an opportunity cost of time when she spends some time explaining the solution to her co-worker.

[11] Of course, it may be that $u_j(h_i = 1) = u_j(h_i = 0)$, for example, when group member j knows how to solve the task in which case $u_j(h_i = 1) = u_j(h_i = 0) = b$.

I conceptualize relative performance concerns as a comparison of performance outcomes.[12] For example, an employee compares her own performance rank $r_i(p_i, p_j)$ to the performance rank $r_j(p_i, p_j)$ of a colleague in her reference group:

$$s(p_i, p_j) = \sigma_t\, f(r_i - r_j)$$

where $f(\cdot): \mathcal{R} \to [0,1]$.[13] The parameter $\sigma_t \in [0,1]$ represents the extent to which relative performance concerns are prevalent in a reference group and the subscript t stands for treatment. I incorporate this into the employee's group morale term:

$$G(a_i, p_i, p_j) = a_i - \sigma_t f(r_i - r_j) \in [-1,1]$$

This captures the idea that an employee's concern for improving the performance of other co-workers is shaped by her relative performance concerns, when these are salient in her reference groups. Initial empirical work on relative performance concerns and unethical behavior in reference groups suggests that $s(p_i, p_j)$ enters the group morale term negatively (Charness, Masclet and Villeval 2014). So does conceptual work on preferences for status. Note that the group morale term could also be negative, in which case an employee may be willing to sabotage her co-worker.[14]

Treatment Manipulations and Group Morale

I hypothesize that relative performance rankings create a sense of competition in reference groups that activates positional concerns over relative performances. Rustichini (2008, p. 653) summarizes the link between competition and relative concerns in a review article on dominance and competition in the following way: "Humans who participate in a contest with others have strong preferences on relative outcomes, and are ready to translate these preferences into costly choices". This implies the following for the salience of relative performance concerns across experimental conditions:

$$\sigma_{baseline} < \sigma_{privateRF} \leq \sigma_{publicRF}$$

The second weak inequality summarizes my hypothesis about public rank feedback. The provision of public performance ranks may further strengthen relative performance concerns since it explicitly invokes social image in intelligence by establishing common knowledge about every group member's performance rank within the reference group.

---

[12] The framework does not explicitly model comparisons of monetary outcomes. Group members did not observe the actual earnings of other group members, since they did not know the absolute performance of others on the timed general knowledge test. Ranks were (weakly) informative about earnings differences in the group.

[13] I do not make any assumptions on the shape of this function $f(\cdot)$ other than bounding the positive or negative impact it can have on the weight that an employee gives to the utility of her co-worker j. In order to detect an effect of relative performance information on group morale in this experiment it just needs to be that $f(x)|_{x \in \{-2,-1,1,2\}} \neq 0$.

[14] The possibility for sabotage is ruled out by design in this experiment.

As the salience of relative performance concerns increases from $\sigma_{t'}$ to σ_t, group morale decreases. To see this, note that in this case, the benefit of helping another group member perform better decreases by $(\sigma_{t'} - \sigma_t)f(r_i - r_j)$, whenever $f(r_i - r_j) \neq 0$.

Thus, in a between-subject design, I can identify the effect of relative performance concerns on the willingness to help others in the reference group perform better by comparing average help in the baseline condition to average help under private or public rank feedback.[15]

A control treatment introduces relative performance pay. Compared to the baseline condition with fixed piece-rates and no rank information, relative performance concerns should be more salient under relative pay, that is, I hypothesize that $\sigma_{baseline} < \sigma_{relative\ pay}$.[16]

In the next section, I present how I manipulated the salience of relative performance concerns across the four experimental conditions and how I measure group morale. Once the design is introduced, I re-state my hypotheses in terms of behavior in the experiment.

3. Experiment Design

3.1.   Overview

In a first part of the experiment, I induce a sense of competition in reference groups by providing relative performance feedback on a task that people perceive as relevant for their social and self-image: a general knowledge test. In a between-subject design, I vary by experimental condition whether or not group members' performances on the test are evaluated relative to one another. In a control treatment, I introduce relative pay that is based on relative performance on this general knowledge test. I then measure, in a second part of the experiment, how the treatment affects the willingness to help others in the reference group perform better. With this design, I test whether a competitive environment in and of itself has a negative effect on group morale. The final parts of the experiment, 3 and 4, elicit further outcome measures and control variables.

Table 1 summarizes the timeline of the experiment. The following paragraphs provide the details for each part.

---

[15] With random assignment to treatment, I can rule out that participants' general inclination to help others, a_i, differs systematically across treatments, that is, with random assignment to treatment it holds that $E[a_i|treatment] = E[a_i]$.

[16] It is not clear whether monetary and non-monetary returns to high relative performance ranks complement or substitute each other. While the salience of relative concerns should not be lower in the relative pay condition, compared to the private rank feedback condition, it need not necessarily be higher.

Table 1. Timeline of Experiment

Part 1 Performance and Feedback
Stage 1 Measure general knowledge
Stage 2 Timed general knowledge test with varying relative performance feedback and pay (by treatment)
Part 2 Measure Group Morale
Part 3 Measure Beliefs
Part 4 Questionnaire

3.2.   The Real Effort Task

Throughout the experiment, the real effort task was to answer multiple choice[17] general knowledge questions for a fix piece-rate.

Participants obtained 25 points for each correct answer. These points were converted at a fixed exchange rate to 1.5 CHF ($\approx$1.5 USD) in pay at the end of the experiment with certainty or with 50% probability whenever the payout of a stage depended on the random draw of a computer.

I chose this real effort task for several reasons. Firstly, participants cared about performing well and being seen as performing well on this task, a fact that was established in pre-tests. This is important since two treatments involved the provision of relative performance feedback on a general knowledge test, without any extrinsic rewards to relative performance.

Secondly, sharing answers to general knowledge questions is a natural way to allow for mutual help on this task. This qualifies as a "real help" task since the outcome of a help decision depends on the performance of the person who provides help and the person who receives it.

Thirdly, performance on this task was not affected by treatment. Participants either knew the answer to these questions or not. Thus, there was little scope for the treatments to systematically change performance outcomes and, consequently, earnings in Part 1. This provides a very clean test of whether a competitive environment in and of itself negatively affects the willingness to help the productivity of others.

Each general knowledge question included in the study was pre-tested in the same subject pool to ensure that the composition of questions in terms of difficulty and the field of general knowledge tested was comparable across parts of the study.[18]

---

[17] Each question had four answer choices.

[18] The objective of the pre-tests was to include questions in the main experiment that would be neither too difficult nor too easy, that questions would be comparable across parts and that there would be no gender differences in performance on average. Average performance data across the different general knowledge tests, i.e., across Part 1 Stages 1 and 2 and Part 2, and by gender show that all of these objectives were fulfilled.

3.3.   Part 1—Performance and Feedback

Part 1 had two stages.

In the first stage of Part 1, every participant was tested on his general knowledge with 10 multiple choice questions. Participants had to provide an answer to each question. Performance on these first questions measures baseline ability at answering the type of general knowledge questions that are used throughout the study. For this first ability measure, everything was held constant across experimental conditions.

At the beginning of the second stage of Part 1, groups were introduced. The computer randomly selected three participants from the same session to form a group. These groups remained fixed for the rest of the experiment. When groups were introduced, each group member saw the portraits of everyone in his group.[19] The pictures were on display for 30 seconds when no instructions were read and there was no option to advance to the next screen. In this way, the timing of when participants saw their other group members for the first time was held constant across all experimental conditions.[20]

In this second stage of Part 1, participants had to answer as many general knowledge questions as possible under some time pressure. This design feature was introduced to ensure heterogeneity in performance on this set of questions.

Participants had 3 minutes to answer a series of 20 multiple choice questions. Questions appeared on their computer screens one at a time and an answer had to be submitted for the next question to appear on the screen. All participants saw the same sequence of the same 20 questions and this known to the participants.

When the three minutes had elapsed, group members were automatically advanced to a feedback screen, whether or not they had provided an answer to all questions. This is the point at which the treatment manipulation occurred, as subjects received different information about their relative performance.

3.4.   Experimental Conditions

When they performed the timed general knowledge test, group members knew what type of feedback they would receive. The feedback screen was on display for one minute and participants were unable to manually advance to the next screen.

In the baseline condition, each group member found out how many out of the 20 questions he answered correctly. No information on the performance of others in the group was provided.

---

[19] Pictures were taken at the beginning of the session by an experimenter.

[20] This is important since Buser and Dreber (2016) find suggestive evidence that a simple group prime may activate norms of competition. In keeping this aspect constant, I rule out that the public rank feedback treatment, which shows subjects pictures of their other two group members, operates through priming the group more strongly compared to the other two conditions.

Thus, group members had no reference point against which to compare their general knowledge score.

In the private rank feedback treatment, a group member also discovered how his performance compared to the performance of others in his reference group. He found out whether he ranked first, second or third in his group. Rankings were based on the number of questions answered correctly during the timed task with ties broken at random. This treatment introduced a performance competition and manipulated self-image in knowledge, relative to group members.

In the public rank feedback treatment, the feedback screen displayed the picture, the participant number and the performance rank of each group member. This way, the relative performance of each individual was common knowledge among the three members of a reference group. Therefore, the public rank feedback treatment made social image in intelligence explicit, while keeping the information about own relative performance the same as in the private rank feedback condition. With this condition, I can assess whether the perceived competition in reference groups is stronger when social image in intelligence is made explicit.

Table 2 summarizes the information shown on the feedback screen in each condition and Figures A1-A3 in Appendix A reproduce images of the feedback screens as they were shown to participants.

A control treatment introduced relative pay in an environment that was otherwise identical to the private rank feedback condition. The best performing group member on the timed general knowledge test received a substantial bonus of 5 CHF ($\approx$5 USD) in addition to the piece-rate that was paid for each correct answer. Thus, the bonus amounted to 25% of the maximum earnings that a group member could receive for this timed general knowledge test. At the feedback stage, a participant observed his own performance rank and whether or not he would receive an additional bonus payment (see Figure A4 in Appendix A). With this condition, I can test to what extent, *ceteris paribus,* any results change with the domain of relative concerns, that is, when money is or is not involved.

Table 2. Experimental Conditions

Baseline	Absolute performance feedback after timed general knowledge test
Private rank feedback treatment	Baseline + private information about own performance rank in group
Public rank feedback treatment	Baseline + public information about everyone's performance rank in group
Relative pay treatment (control)	Private rank feedback + best performer on timed test earns an additional bonus of 5 CHF

3.5.   Part 2—Measuring Group Morale

The feedback that subjects saw at the end of Part 1 was also summarized in their Part 2 decision screens.

In Part 2, group morale was measured in a task that was independent of the relative performance competition in Part 1. Part 2 comprised 10 new multiple choice general knowledge questions. Each group member had to provide an answer to each question. The piece-rate for correct answers stayed the same as in Part 1, that is, 25 points or 1.5 CHF.

For each question in Part 2, a participant had the option to share his answer with the other two group members. I chose this type of task explicitly to model the kind of helping behavior that takes place in workplace settings, where someone who knows information (how to accomplish a task, the needs of a particular client) can share this information with others to help their productivity.

Whenever a participant shared a correct answer to a question, the computer automatically replaced the incorrect answer of each group member who did not answer that question right with the correct answer that was shared. Sharing incorrect answers had no positive or negative effect on others in the group. This way, I ruled out by design that participants could sabotage the performance of others by sharing incorrect knowledge with them. The main reason is that, in organizations, on-the-job help is typically a prosocial task and help behavior in the experiment reflects this.

The total benefit to sharing an answer was either 0, 25 or 50 points (0, 1.5 or 3 USD)—depending on whether a correct answer was shared and on how many group members did not get a question right. This benefit went to *others* in the group.

Sharing an answer to a question cost a participant 1 point or 0.06 CHF ($\approx$0.06 USD). Thus, when a group member shared an answer he was willing to invest 1 point in the performance of others. That amounts to 4% of the piece-rate paid for correct answers.[21] This small cost to knowledge sharing mirrors that in real organizations, on-the-job help comes with an opportunity cost. Moreover, that helping behavior in the experiment is, unambiguously, a prosocial act.

This way, I elicited participants' willingness to help others to perform better, observing 10 decisions for each person regarding whether or not to share the answer to a Part 2 question with others in the group. Figure A5 in Appendix A shows a screenshot of the decision screen. All instructions for participants described the act of sharing answers with the more neutral term "sending answers" to other group members.

---

[21] This piece-rate was calibrated with a pilot study of the baseline condition to ensure that the level of knowledge-sharing in the baseline condition was neither too high (i.e. above 75% or higher) or to low (i.e. below 25%). In a first calibration of the design, the personal cost to sharing answers with others was 10% of the piece-rate under which the level of knowledge-sharing was too low.

Participants did not obtain feedback about choices that others made in Part 2. Participants were aware of this when they made their help decisions. Importantly, participants did not see who helped them, how much help they received or how many of their answers were replaced. I gave no feedback on help decisions to rule out that group members could seek to enhance their social image in a second dimension, namely, a reputation for being prosocial. This could interact with the treatments making it more difficult to isolate the direct effect of relative performance evaluations on the willingness to invest in the productivity of others. At the very end of the experiment, participants found out how many Part-2-questions they answered correctly when they saw their summary of earnings in the experiment.

When participants performed in Part 1 they did not know anything specific about the other parts of the experiment. Importantly, participants were unaware that Part 2 would entail prosocial choices. This allows me to rule out that any effect of rank feedback on the willingness to help other group members is driven by selection into high rank positions based on social preferences, similarly to what Erkal, Gangadharan and Nikiforakis (2011) find in a study on competing for money and the subsequent willingness to redistribute earnings within a group.

3.6.   Part 3—Beliefs

In Part 3 of the experiment, I elicited three beliefs for each Part-2-question; the subjective probabilities that a participant assigned to his answer and to the answer of each of the other two group members being correct.

I implemented a mechanism to elicit subjective probabilities in an incentive compatible way that was described in Karni (2009) closely following the experimental protocol introduced in Coffman (2014). In this part of the experiment, there were 100 lotteries available that had an integer-probability on [1,100] of selecting a correct answer to a question. In other words, there were lotteries that had a 1% chance, 2% chance, 3% chance ... up to a 100% chance to provide a correct answer to a question. For each question, one of these lotteries was randomly selected, with each lottery equally likely to be chosen. Participants selected a threshold, X, such that for any lottery that selects the correct answer with a probability X or lower they would prefer their own answer to be evaluated for payment and for all lotteries that select the correct answer with a probability X or higher they would prefer the lottery to answer for them. Therefore, given a cut-off probability X', a participant believes that the answer he provided to a question is correct with probability X'.

For each question, participants stated three different such cut-off probabilities: one for the answer they provided themselves and one for the answer provided by each of the other group members. For payment, one of each of the three "types" of belief (self, two other group members) was randomly selected and evaluated. In this part, participants earned 2 CHF ($\approx$2 USD) if they

submitted a correct answer, regardless of whether this answer was submitted by them, by one of their group members or by a lottery.

For this belief elicitation task, the order of Part-2 questions was randomized at the subject-level.

With this data, I can assess whether participants intended to use help decisions instrumentally, to help group members. Sharing answers that one believes to not know is not helpful, neither is it to share answers that one believes the others to know for sure. Moreover, I can use this data to account for the pure information effect of rank feedback through which the treatment may systematically affect the willingness to help compared to the baseline with no relative performance information.

I also asked participants to state how much help they expected to have received from each of the other two participants in their group. Participants earned 1 CHF when their guess of the number of answers that a group member shared with the group was within a margin of +/- 1 question to the actual number of questions this person sent. Expectations about others' help decisions could also influence the willingness to share knowledge, out of a desire to reciprocate or a desire to conform. Since rank feedback could also influence the help a group member expects to receive from others, it is important to have data on expected help.

3.7.　Part 4—Questionnaire

In an exit questionnaire, I collected several measures to assess perceptions of social relations in the experimental reference groups. I measured the intensity of perceived competition in the experimental (reference) groups with an agreement to the following statement: "I felt in competition with the other two members in my group when performing this task." on a 9-point Likert scale. "This task" refers to the timed general knowledge test of Part 1. This data was collected about 30 minutes after this stage of the experiment and serves as a test of whether relative performance feedback produced a sense of competition among the members of the reference group.

On the same scale, participants also answered questions to assess to what extent performing well on the general knowledge test and being seen performing well by others was desirable and to evaluate whether they thought that the questions actually tested general knowledge.

In addition, I measured the social distance among members of experimental reference groups with the Oneness index. It records the subjective perception of closeness between a participant and each of his group members, ranging from "no connection at all" to feeling "at one" with another person. This scale is widely used in psychology to measure the closeness of social relationships (Gächter, Starmer, and Tufano 2015a) and predicts behavior in economics studies involving decision-making in groups (see, e.g., Gächter, Starmer, and Tufano 2015b).

The Oneness index is constructed from responses to the Inclusion of the Self in Other (IOS)-scale and the We-scale. On the IOS-scale, a participant indicated how close he felt to another group member by selecting a pair of circles that best represents the relationship with that other group member. In these pairs, one circle depicts the participant and the other circle the other group member. Across the pairs, the circles differ in how much they overlap. See Figure A6 in Appendix A for the pictogram used in this task. For the We-scale, a participant indicated on a 7-point Likert scale to what extent he would use the term "we" to characterize himself and another group member. The Oneness index is simply the average of a participant's responses on these two scales for a given group member. As a measure of social relations in reference groups, this outcome variable provides complementary evidence for the change in competitiveness from the baseline to the treatment conditions.

To complement my behavioral measure of group morale, I also elicited general attitudes toward cooperation, toward working in groups or working alone and toward competition, following the procedure for eliciting general attitudes presented in Duffy and Kornienko (2010). I also took their set of items eliciting attitudes on competition. For each of these general attitudes, participants evaluated the extent to which four statements applied to them on a 9-point Likert scale ranging from 1 -does not apply at all, to 9 -definitely applies.[22] For each category, e.g. for cooperativeness, an index is constructed which is the average response to the four statements that belong to the category, reverse-scoring responses when necessary. A list of all four items for each category with summary statistics is in Appendix A (Table A2).

Lastly, I elicited positive and negative reciprocity as general traits with the set of questions described in Falk et al. (2016). The authors provide evidence that responses to these questions are highly correlated with behavior in experimental games that are typically used to measure reciprocity with decisions involving real monetary stakes. This allows me to test whether the treatment—that may impact prosocial help—also extends to general prosocial attitudes.

The questionnaire concluded with a few questions on demographics and an elicitation of attitudes towards risk and towards competition. For each of these attitude measures, participants were asked to position themselves on a scale from 0 (very risk-averse; not competitive at all) to 10 (very risk-seeking; very competitive).

3.8.   Hypotheses About Behavior in the Experiment

The experiment is designed to assess whether relative performance information lower group morale when participants make 10 help decisions in Part 2. I hypothesize that relative performance evaluations put group members in a competitive mindset with respect to one another, compared to an environment that, *ceteris paribus*, does not provide this information:

---

[22] For example, participants rated "I like to share my ideas and material with others." which is a statement on cooperation or "I find that working in groups is often inefficient" which is an attitude toward working in groups. I included four questions per attitude to reduce the influence of an idiosyncratic question.

Hypothesis 1: Rank feedback causes a sense of competition in the reference group in Part 1.

I test the mechanism that this competition activates positional concerns over relative performances, which may lower the motivation to subsequently extend prosocial help to others in order to increase their productivity (see also section 2). My study offers a direct test of the null hypothesis,

Hypothesis 2-0: The sense of competition under relative performance feedback does not lower the willingness to help other group members perform better in Part 2.

against the alternative hypothesis that,

Hypothesis 2-A: The sense of competition under relative performance feedback lowers the willingness to help other group members perform better in Part 2.

3.9. Experimental Procedures

The experiment was conducted in English at the Laboratory for Experimental and Behavioral Economics at the University of Zurich. In total, 282 participants, most of them students at the University of Zurich and the Swiss Federal Institute of Technology in Zurich, took part in the experiment. Table 3 lists the number of participants in the different conditions.[23]

At the beginning of a session, an experimenter took pictures of all participants before participants took their seat in the laboratory. Participants were called individually by their participant number and were instructed to make a neutral face for the portrait. The composition was the same for every portrait, with a zoom on the face leaving out the upper body. Participants gave informed consent to having their picture taken and to the fact that these pictures may be linked to some of their choices in the experiment.

Table 3. Overview of Data

Condition	Participants
Baseline	obs.=72, 24 groups
Private rank feedback	obs.=72, 24 groups
Public rank feedback	obs.=66, 22 groups
Relative Pay	obs.=72, 24 groups
Total	obs.=282, 94 groups

---

[23] While 7 out of 9 sessions comprised 24 participants (8 groups), two sessions in the public rank feedback condition were conducted with 21 participants (7 groups) because some of those who registered did not show up.

The instructions for the study were displayed on the computer screen in a participant's cubicle. Screenshots of the instructions and decision screens exactly as they were shown to participants are reproduced in Appendix B. An experimenter read the instructions for a part out loud just before participants made choices in that part. Before Part 2 and Part 3, participants also answered comprehension questions and the study only advanced after all participants had answered the questions correctly.

The experiment was programmed in z-Tree (Fischbacher 2007). The computer selected for each participant whether the first stage of Part 1 or the second stage of Part 1 was selected for payment, giving equal weight to each option. Earnings in Parts 2 and 3 were always paid out. Average earnings were 40.00 CHF ($\approx$40 USD) (including a 15 CHF show-up fee).

4. Results

First, I will consider results in support of Hypothesis 1. Then, I will turn to results on the treatment effect of relative performance information on group morale that lead to the main result of the paper regarding Hypotheses 2-0 and 2-A. I will then briefly present findings on the factors that predict helping behavior and consider findings from a control treatment that introduces relative pay. I conclude this section with results on the link between group morale and group productivity in the environment under study.

4.1. Do Relative Performance Evaluations Affect Perceptions of Competition?

Rank feedback on the timed general knowledge test of Part 1 mirrored actual performance differences among the group members in the absolute majority of reference groups.

Performance on the timed general knowledge test of Part 1 varied substantially and performance ties occurred only in 12 % of the randomly formed experimental groups. On average, participants answered 11.3 out of 20 questions correctly (SD=3.0). The best performers in my sample answered 17 questions correctly and the worst performers 3. The empirical distributions of performance on this timed task are very similar across the four experimental conditions (see Figure A7 in Appendix A) and a Kruskal-Wallis test fails to reject the null hypothesis that these performance samples are drawn from the same population (p=0.4346).[24] This is expected given that the task tested existing knowledge and effort had little to no scope to increase performance on this test.

---

[24] Consistent with this, participants did also not perform systematically differently across the four experimental conditions on the first set of questions that was administered before the treatment manipulation occurred: Kruksal-Wallis test p-value=0.5410. On average, participants provided correct answers to 61% of questions on this first untimed general knowledge test (mean number of questions correct=6.160, SD=1.640).

Hypothesis 1 states that relative evaluations on this first timed general knowledge test produces perceptions of competition between group members. To address this question, I turn to data from the post-treatment questionnaire, in which several items asked specifically about the timed task in Part 1. I focus on the causal effect of providing relative performance feedback on participants' agreement with the statement "I felt in competition with the other two members in my group when performing the task." Responses can range from 1 (does not apply at all) to 9 (definitely applies). Figure 1 visualizes the marked differences in the perception of competition in groups between the baseline and the two rank feedback conditions. In the baseline condition, the average sense of competition in groups is 2.0 on this competitiveness scale which is very low. The estimated average treatment effect of private rank feedback is an increase of 3.4 points on the 9-point competitiveness scale with a 95% confidence interval of [2.68, 4.12]. The estimated effect of public rank feedback is a 3.8-point increase in the sense of competition with others in the reference group with a 95% confidence interval of [3.10, 4.49]. These effects are highly statistically significant (two-sided t-tests: Bas-Priv. RF $p<0.0001$; Bas-Pub. RF $p< 0.0001$). The fact that this data was collected 30 minutes after participants completed Part 1 indicates that this information had a lasting impact on the sense of competition in reference groups.

Whether the performance feedback is public or private appears to not make a sizable difference in the sense of competition in reference groups. The effect of a public ranking over a private ranking is 0.40 points on the competitiveness scale with a 95% confidence interval of [-0.51, 1.29], which includes 0 and is not statistically significant (two-sided t-test $p=0.3926$).

Importantly, while relative performance feedback changed the perception of competition among group members, other aspects of how individuals evaluated the task in Part 1 did not change. Using a 9-point scale on which higher numbers indicate more agreement with a statement, participants thought that the general knowledge questions did, in fact, measure their general knowledge (mean agreement=7.26, SD=1.83). Participants also wanted to perform well at the Part 1 performance stage (mean agreement=8.15, SD=1.32) and would be impressed if others answered 90% or more of the general knowledge questions in the experiment correctly (mean agreement=7.35, SD=2.12). The treatments had virtually no effect on this (see Figure A8 in Appendix A for a visualization). All in all, this supports the interpretation that the timed task of Part 1 is relevant for perceptions of competence, and that individuals valued performing well and being seen as performing well on the task.

Figure 1. Perceptions of Competition by Condition
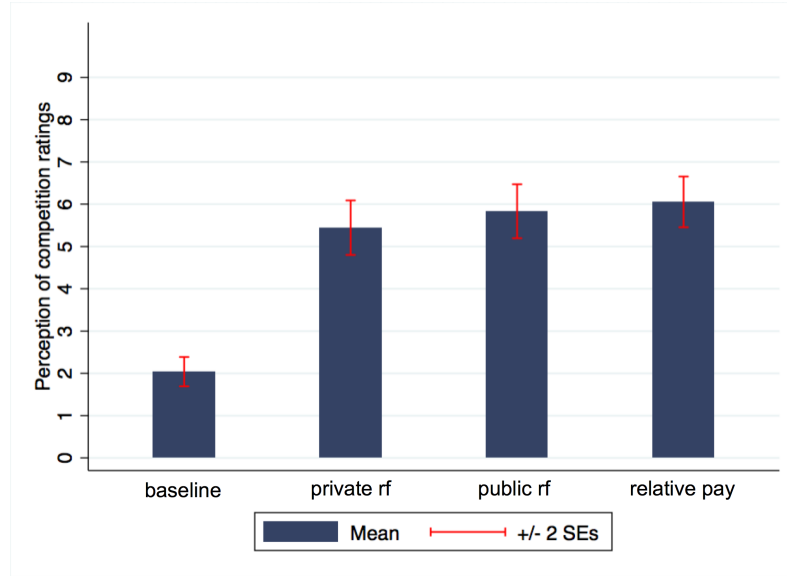
Note. Perception of competition is measured as agreement with the statement "I felt in competition with the other two members in my group when performing this task [Part 1 timed general knowledge test]." on a scale from 1 ("does not apply at all") to 9 ("definitely applies").

Relative performance evaluations also increase the perceived social distance between members of experimental reference groups. The Oneness index (Gächter, Starmer, and Tufano 2015b) is a simple instrument to measure how close subjects perceive themselves to be to every other group member on a scale from 1 (no connection at all) to 7 (feeling as "one" with another person). In column 1 of Table 4, I present the results from ordinary least square (OLS) regressions of the Oneness index, averaged over the two group members of a participant, on treatment indicators. The results indicate that rank feedback decreased the social closeness between members of experimental reference groups, when group members were asked 35 minutes after the end of Part 1 how close they felt to others in their group. The estimated average treatment effect is 0.51, or half a category on this 7-point scale, under private rank information and 0.40 under public rank information.

In contrast, *general* attitudes toward working in groups or working alone were not systematically affected by the treatments (columns 2) and 3) of Table 4).

Table 4. Rank Feedback, Relative Pay, Social Closeness and General Attitudes on Groupwork

	Social closeness with group members	General attitude working in groups	General attitude working alone
Private RF	-0.514***	-0.243	0.191
	(0.181)	(0.229)	(0.187)
Public RF	-0.399*	0.155	0.071
	(0.206)	(0.215)	(0.200)
Relative Pay	-0.469**	-0.215	0.184
	(0.205)	(0.234)	(0.230)
Constant	3.028***	5.083***	6.038***
	(0.130)	(0.154)	(0.122)
Obs.	282	282	282

Notes. The data was collected for each participant i in a questionnaire at the end of the study. *Private RF* and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Relative Pay* indicates that participant i was in the control treatment that paid a bonus to the best performer, in addition to providing rank feedback in private. Social closeness with group members is the average of the two responses of i on the oneness index measuring how close i feels to each group member j. This variable ranges from [0,7]. General attitudes on working in groups and on working alone are indices that range from 1 (strongly negative attitude) to 9 (strongly positive attitude). Robust standard errors are in parentheses, 94 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level.

All these findings support the conclusion that relative performance evaluations changed the perceptions of social relations within the experimental reference groups. This leads to the following first result:

Result 1: Rank feedback information causes perceptions of competition between group members and increases their social distance.

How does this increase in the salience of relative performance concerns under competition affect the willingness of group members to help each other perform better? These are the next results I turn to.

4.2.   Do Increased Perceptions of Competition Affect Group Morale?

Statistical Hypothesis Testing

In the following section, I present results from analyzing help behavior in Part 2 of the experiment with a focus on the treatment effect of relative performance feedback. Throughout this section, I compare behavior in the baseline condition to behavior in the two treatments that

provided relative performance information, that is, the private rank feedback and public rank feedback treatment conditions.

First, I estimate the causal effect of relative performance evaluations on the total number of answers a group member shared out of the 10 Part 2 questions (see Table 5).

Table 5. Average Treatment Effect of Rank Feedback on Help

ATE $\mu_{bas} - \mu_{treat}$	95% confidence interval
Bas-Priv. RF: 0.36	[-0.69, 1.41]
Bas-Pub. RF: 0.013	[-1.12,1.15]

The estimated average treatment effect of private rank feedback is a decrease in 0.36 questions shared, with a 95% confidence interval of [-0.69,1.41] and a p-value of 0.50. This confidence interval includes zero and the decrease in average help of 3.6% is not statistically significant at any reasonable significance threshold. The estimated average treatment effect of relative performance feedback is even smaller, a decrease in 0.013 questions shared with a confidence interval of [-1.12, 1.15] that includes zero and a p-value of 0.98.[25] It is important to highlight that the estimated confidence intervals are narrow, which means that the average treatment effects are precisely estimated. Section 6 gives a detailed account of the statistical power of this study.

Figure 2 shows that also the empirical distributions of number of answers shared look very similar across the three experimental conditions. A Wilcoxon-Mann-Whitney test fails to reject the null hypothesis that the baseline and treatment samples of help behavior are drawn from the same population (Bas-Priv. RF p=0.535, Bas-Pub. RF p=0.775).

The empirical distributions of help also visualize that there are three different empirical "types" of people. The "resolute helpers" share all ten answers to their questions and the "selfish"-type does not share any answers. The majority of people share some answers with others. In the baseline condition, for example, 11% are resolute helpers and 17% of group members are selfish.

The shares of the selfish-type and the resolute helpers are remarkably consistent across the three experimental conditions (see Figure 2). This is further evidence suggesting that the treatments did not systematically affect behavioral motives for knowledge-sharing.

---

[25] These are the Average Treatment Effects (ATE) and confidence intervals when I treat each group member as one observation. The precision of my ATE estimates are very similar when I collapse observations at the group level, which takes into account that group members may have been exposed to some common shocks. The 95%-confidence intervals are Bas-Priv. RF [-0.65,1.37] with a p-value of 0.477 and Bas-Pub. RF [-1.04,1.07] with a p-value of 0.981.

Figure 2. Empirical CDF of Help by Condition
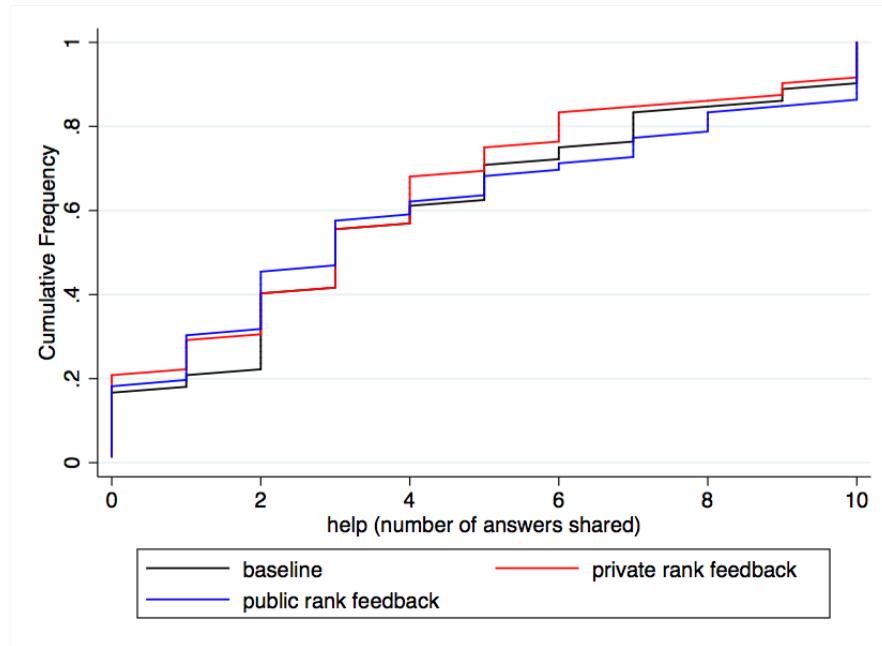
So far, there is no strong evidence in support of hypothesis H2-A, that relative performance concerns, activated by relative performance evaluations, have a sizable negative effect on group morale.

However, with this type of statistical hypothesis testing one cannot control for the fact that beliefs about the value of own help to the group may be affected by relative performance information, which may counteract a negative treatment effect.

Relative Performance Information and Confidence in Self and Others

I find no evidence that self-confidence in answering general knowledge questions correctly is affected by rank feedback information.[26] But I find that knowing performance ranks has a small but measurable impact on the confidence in the ability of the other two group members to answer general knowledge questions correctly in the expected direction.[27] Taken together, these results imply that, for example, rank 1 group members who know their rank deem their help to be slightly more valuable, on average, than rank 1 group members in the baseline condition. This can potentially counteract a negative effect of relative performance concerns.

---

[26] Table B1 in Appendix B reports the results from OLS regressions that predict, conditional on performance rank, the percent chance that a participant assigns to himself having provided a correct answer to a Part 2 question as a function of treatment indicators and controls.

[27] Table A1 in Appendix A presents results from OLS regressions that predict the percent chance that a participant assigns to another group member providing a correct answer to a Part 2 question, conditional on the rank of the participant who makes this judgement, treatment indicators and further control variables. The results indicate that relative performance information systematically affects the confidence in other group members' likelihood of providing correct answers in the expected direction, for rank 1, rank 2 and rank 3 group members in the private and the public rank feedback conditions. For example, observing their relative performance rank in private (public) reduces rank 1 group members' confidence in their other group members' ability to provide correct answers to Part 2 questions by about 5.5 (7.2) percentage points.

Binary Choice Models Predicting the Willingness to Help

In binary choice models that predict the willingness to share answers with others, I can control for beliefs about correct answers. I estimated the following general model:

$$Prob(share_{ki}) = g\big(\beta_0 + \beta_1 I_{privRF} + \beta_2 I_{pubRF} + \beta_3 belief\ self\ correct \\ + \beta_4 belief\ others\ in\ group\ correct + [Controls_i + FE_k]\big)$$

share_{ki} indicates whether participant *i* shared his answer to question *k* with the others. I_{privRF}, I_{pubRF} indicate whether this participant privately observed his performance rank on the timed task or publicly observed the performance rank of everyone in his group. The variable *belief correct self* is the subjective probability that group member *i* gave to his answer to question *k* being correct. The variable *belief others in group correct* is the average of the probabilities that group member *i* gave to his two group members having provided a correct answer to question *k*. I introduce the other control variables below when I briefly consider results on the factors that correlate with helping decisions (section 5.3). I report results from Probit and Linear Probability Models.

Table 6 columns 1)-3) show the results from fitting a Probit model. The first specification predicts the willingness to help as a function of treatment indicators with no further control variables. The predicted marginal effect of private rank feedback or public rank feedback on the willingness to help are very small and statistically insignificant (private RF: -0.036, public RF: -0.001) which confirms the previous findings from parametric and non-parametric hypothesis testing.

The second specification adds controls for beliefs about correct answers. The willingness to share the answer to a question increases in the probability that a group member assigns to his answer being correct (*belief correct (self),* predicted marginal effect at means of covariates=66 percentage points) and decreases as he gives a higher probability to his group members having provided a correct answer (*belief correct (others),* predicted marginal effect at means of covariates=-28 percentage points). Both effects are highly statistically significant. The signs of the two coefficients, positive for *belief correct (self)* and negative for *belief correct (others)*, indicate that participants help more when they believe their help will be valuable. [28]

Importantly, however, the introduction of these control variables does not change the inference about the treatment effect of relative performance evaluations on the willingness to

---

[28]In Appendix B, I also present results from models in which I relate these belief variables, instead of testing for them separately (Table B3). I construct a control variable *valuable help* that ranges from [0,2]. It sums the two conditional probabilities that a participant *i* assigns to her answer improving the performance of group member 1 and group member 2, conditional on the group member not knowing the answer to that question. I can back out these probabilities from the three beliefs about correct answers that participant *i* stated for each Part 2 question under the assumption that they are independent. None of the results presented in the main text change when I control for the expected value of own help to others in this way.

help others perform better, the predicted marginal effects remain very close to zero (private RF -0.038, public RF 0.006).

Table B2 in Appendix B presents results from specifications similar to the ones presented in columns 1-3 of Table 6 but pooling the data from the private and public rank feedback treatments. This further increases the power to detect a treatment effect. I pool the data from these two treatments because they had very comparable effects on the perceptions of competition in Part 1 (see Figure 1). These results also fail to reject hypothesis H2-0.

I also perform additional robustness analysis in which I add terms that interact each treatment indicator with the variables *belief correct (self)* and *belief correct (others)* in a linear probability model. This directly tests whether participants are less likely to share the answers that they think are going to be valuable to others under either type of relative performance feedback. While the estimated sign of the interaction terms is typically negative—consistent with hypothesis H2-A—they are far from reaching statistical significance (see Table B4 in Appendix B).

Heterogeneity in Treatment Effect by Rank in Competition

Looking at average treatment effects in the whole sample could mask substantial heterogeneity in how strongly group members of different performance ranks change their help behavior after rank feedback. **Error! Reference source not found.** displays average helping by rank and by condition. For rank 1 and rank 2 group members, average help is slightly lower under relative performance evaluations, whereas it is slightly higher for rank 3 group members.

Figure 3. Help by Treatment and Rank in Competition (Means)
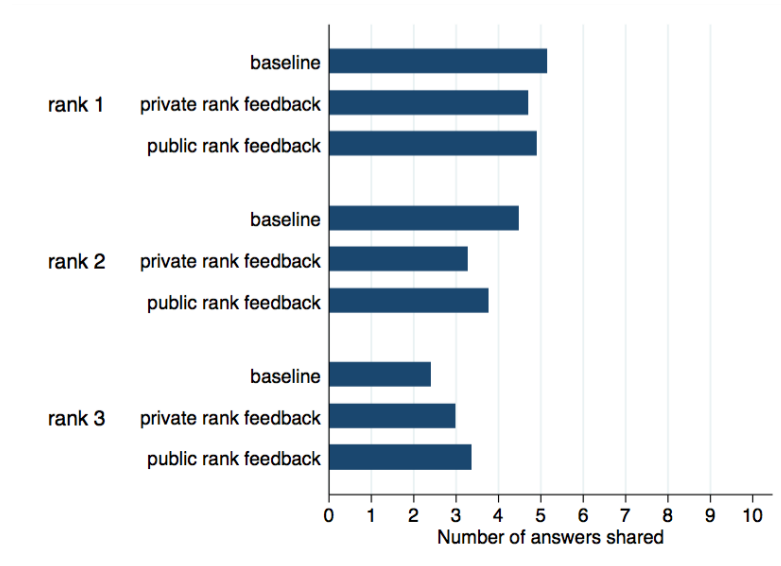
Table 6. Predicting the Willingness to Help

	Predicting Prob(share answer to question k)			
	Probit			OLS
Private RF	-0.03623	-0.0388	-0.0132	-0.00560
	(0.0500)	(0.0562)	(0.0485)	(0.0330)
Public RF	-0.0012	0.006	0.0366	0.0230
	(0.0522)	(0.0571)	(0.0458)	(0.0310)
Belief correct (self)		0.659****	0.7803****	0.558****
		(0.0947)	(0.1314)	(0.0654)
Belief correct (others)		-0.2762**	-0.3682***	-0.269****
		(0.1126)	(0.1404)	(0.0767)
Actual correct			-0.0088	-0.00856
			(0.0342)	(0.0225)
Performance part 1			0.1563	0.125
			(0.1506)	(0.0998)
Expected help			1.4436****	1.053****
			(0.1776)	(0.0661)
Risk attitude			-0.2478****	-0.150***
			(0.0768)	(0.0552)
Female			-0.0294	-0.0234
			(0.0487)	(0.0340)
Question FE	No	No	Yes	Yes
Constant	Yes	Yes	Yes	-0.334****
				(0.0892)
Obs.	2100	2100	2100	2100
(pseudo) R^2	0.0009	0.08	0.35	0.381

Notes. Predicted marginal effects at mean level of covariates are reported (Probit). *Share answer to question k* is an indicator for whether participant shared the answer to a Part 2 question k with others. *Private RF* and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in the group. *Belief correct self* ranges from [0,1] and is the subjective probability that participant i gives to the event that his answer to question k is correct. *Belief correct others* ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to question k. Actual correct indicates whether participant i provided the correct answer to question k. *Performance Part 1* records the fraction of questions that participant i answered correctly during the timed task of Part 1. *Expected help* is the fraction of answers hat participant i thinks the two group members shared, on average. *Risk attitude* is where participant i positioned himself on a scale from 0=very risk-averse to 10=very risk seeking, divided by 10. Robust standard errors are in parentheses, 70 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

However, Wilcoxon-Mann-Whitney tests fail to reject the null hypothesis that there are any differences in how help, conditional on rank, is distributed comparing help behavior in the baseline condition with help behavior under either type of relative performance feedback (see Table A5 in Appendix A for a summary of p-values).[29]

Within each condition, higher ranked group members typically provide more help, on average, than lower ranked group members, consistent with an interpretation that participants use the help technology because they seek to improve the performance of others in their group.Finally, there is another way to consider heterogeneity in treatment effects based on performance on the timed task of Part 1. Note that relative performance information could be particularly relevant for individuals who perform neither exceptionally well nor very poorly on the timed general knowledge test. The absolute performance feedback in the baseline condition may have provided participants whose scores were on the tails of the performance distribution, e.g. 4 questions or 16 questions answered correctly, already with a clear sense of how these outcomes compare to the performance of others, thereby activating relative performance concerns also in the baseline. In additional robustness analysis, I account for the fact that the value of the treatment information may differ according to one's absolute performance on the timed general knowledge task. Table B6 in Appendix B presents results of a Probit estimation that predicts the probability to share answers for subsamples of participants whose performance on the timed general knowledge test was neither exceptionally good nor bad according to performance percentiles.

This additional robustness analysis confirms the finding that relative performance information *per se* appears to not lower the intrinsic motivation to help others perform better.

This leads to the main result of the paper:

Result 2: Hypothesis 2-0, that the performance competition under relative performance feedback has no effect on the willingness to help other group members perform better, cannot be rejected.

---

[29] In Appendix B I report results from binary choice models that predict the willingness to help conditional on performance rank, treatment indicators and controls for beliefs about correct answers. For this analysis, the data from the private rank feedback and public rank feedback treatments are pooled (Table B5). Even when I pool the data, for this type of analysis, the sample size is relatively small. This makes the inference that one can draw based on this sample limited, though the relatively small magnitudes suggest modest effects of the treatment on the willingness to help also when I allow for heterogeneous treatment effects. The point estimates of the treatment effect tend to be small in magnitude, e.g. the linear probability model predicts a marginal effect on the willingness to help by 6-11 percentage points, and are never statistically significant.

Relative Performance Evaluations, Expected Help and Cooperative Behavior in General

Consistent with this main result, the treatment manipulations do also not extend to have effects on neither the expected help behavior of others nor the general willingness to act cooperatively (see the results in Tabe A3 in Appendix A).

To summarize, while the private and public rank feedback treatments reliably change perceptions of social relations in reference groups, that is, the sense of competition and social distance between group members, they do not systematically affect actual help behavior, expectations about the help behavior by other group members, the desire to help others in general or broader prosocial inclinations.

4.3.   What Predicts Variation in Help?

In the following section, I consider variables other than treatment status and how they contribute to explaining variation in the willingness to help. This analysis provides insights into what motivates participants to share their answers with others.

Column 3 of Table 6 shows results of a Probit estimation that predicts the willingness to help as a function of treatment indicators, performance controls, beliefs about correct answers, the expected help by others, risk attitudes, gender and question fixed effects.

We have already seen that *beliefs about correct answers* are highly predictive of helping decisions. This suggests that participants share answers that they think will help other group members' performance. This lends support to the interpretation that the task in Part 2 elicits group morale, that is, subjects share knowledge in order to improve the performance of others.

The variable *expected help* ranges from [0,1] and is the fraction of Part 2 answers that a participant expected to have received from the two group members, on average. The point estimates of the coefficient suggest that the expected help by others has a sizeable effect on the willingness to help them perform better. [30] This is consistent with the interpretation that behavior in Part 2 is motivated by prosocial inclinations, i.e. positive reciprocity in the form of mutual positive expectations in a group. [31] This provides further evidence suggesting that the task in Part 2 operationalizes group morale, whereby group members are more willing to help others who they think will also help them.

The coefficient on the measure of *risk attitude* suggests that individuals who describe themselves as risk-seeking are less willing to share their answers with others than individuals

---

[30] Note that the estimated and predicted marginal effect of the variable "Expected help" *at the mean level of other covariates* exceeds one. The predicted *average marginal effect* of the variable "Expected help" is, of course, smaller than one (0.96025 with a standard error of 0.0669).

[31] An alternative interpretation is more mechanical, whereby participants base expectations of others' help on their own behavior.

who describe themselves as risk-averse. This may indicate that one motive behind knowledge-sharing is to insure other group members against a lack of knowledge on a particular question.

Conditional on this full set of controls, there are no *gender* differences in the willingness to share answers with others. In this context, women do not act systematically more prosocial than men.

Lastly, there is no association between *absolute performance* on the timed general knowledge test and the willingness to help others in Part 2 of the experiment. In this decision context, the higher experimental earnings that group members were aware of when making help decisions did not translate into an increase in generosity.

The corresponding OLS estimates, that largely agree with the Probit estimates, are reported in column 4 of Table 6.

4.4.   Relative Pay Treatment

Next, I turn to the results from a control treatment that entails relative pay.[32] With this data, I can directly compare the intensity of competition under relative performance information to the benchmark of a money competition.

Competing for relative pay has a large and significant increase (about 4-points on a 9-point scale) on the sense of competition in reference groups compared to the baseline in which relative concerns are absent (see Figure 1).

I find no evidence that competing for relative pay increases the intensity of competition in reference groups, compared to otherwise identical environments in which group members only obtain relative performance information in private (two-sided t-test, p-value=0.166) or public (two-sided t-test, p-value=0.612).

Results from an OLS regression indicate that relative pay also increases the social distance among members of a reference group (see Table 4 column 1). Competing for relative pay reduces the reported social closeness to the other group members by -0.469 points on average. This negative effect on the perceived social closeness among members of a reference group is comparable, in terms of absolute effect size, to how harmful the provision of private and public relative performance information is (see section 5.1).

Taken together, these findings support the interpretation that the intensity of competition in reference groups is comparable across the domains of relative performance comparisons and relative pay.

---

[32] This treatment is the same as the private relative performance feedback treatment in terms of relative performance information that group members obtain at the end of Part 1 of the experiment. The best performer of a group makes a substantial monetary bonus that amounts to 25% od potential Part 1 earnings.

Consistent with this first result, I also find no evidence that competing for relative pay has a sizable negative (or positive[33]) effect on the willingness to extend costly help to improve the performance of other group members. The average treatment effect of relative pay on help is 0.639 answers shared less after the experience of a competition for pay, with a 95%-confidence interval [-0.45, 1.72] that includes zero. Results from choice models that predict the willingness to help others perform better as a function of an indicator variable for relative pay and controls for beliefs about correct answers confirm this conclusion (Table A4 in Appendix A).

These findings provide further evidence that a competitive environment in and of itself does not systematically affect the willingness to extend costly help to improve the productivity of others.

4.5.   Missed Opportunities to Help and Productivity

This paper sets out with the observation that it is important to understand what makes groups productive, in order to obtain a better understanding of the factors that determine productivity in firms. Next, I zoom in on the counterfactuals that are seldom observed outside of the laboratory. I identify the missed opportunities to help other group members perform better and quantify what they imply for group productivity and efficiency.

A missed opportunity is defined as an instance in which one group member submitted the correct answer to a question but did not share the knowledge, with the result that another group member ended up submitting the wrong answer to a question. Each group in my experiment had, on average, 4.0 of such missed opportunities (SD=2.7).[34] Group productivity would have increased by 18%, on average, had these answers been shared. The average efficiency loss of missed opportunities—unrealized group earnings under higher productivity net the cost of help— is 5.8 CHF (≈5.8 USD) (SD=4.0). Table A6 in Appendix A also reports these numbers separately for each experimental condition.

This analysis shows that a change in group morale, a decrease or increase, would have real consequences on group productivity in the environment under study.

5. Discussion & Concluding Remarks

The study investigates whether relative performance evaluations lower group morale in reference groups. In a laboratory experiment, I vary by treatment whether or not members of a

---

[33] In the relative pay treatment, the Part 1 earnings of the best performer are at least 25% higher compared to the other two group members. I find no evidence that this has any effect on the winner's generosity to extend costly help, compared to the best performers in the baseline condition (ranskum test p-value=0.9583).

[34] Group members who did not share their answers, although they got it right and sharing could have improved the performance of at least one group member, had an average confidence of 64% that their answer was correct; in 50% of these cases group members gave a chance of at least 70% to their answer being correct. This indicates that the majority of these missed opportunities are, indeed, missed opportunities rather than people having simply guessed the right answer to a question that they did not share because they thought they would not know the answer.

reference group obtain relative performance feedback on a timed general knowledge test. Rank feedback is either given in private or in public. I then test how the salience of relative performance concerns under relative performance feedback spills over to the willingness to help others in the reference group perform better.

The data show that relative performance evaluations on a task that people perceive as relevant for their social and self-image cause a large and lasting increase in the sense of competition in reference groups, compared to the baseline condition in which no relative performance feedback is provided. The intensity of competition under private or public rank feedback is comparable to how strongly members of reference groups compete for relative pay in a control treatment.

Yet, I find no evidence that relative performance concerns from the competition spill over to subsequent help behavior. Relative performance evaluations and relative pay do not systematically change the willingness to help others in the group. Beliefs about the value of own help in terms of improving other group members' performance are a strong predictor of actual help behavior, which is consistent with the interpretation that participants share knowledge in order to help the performance of others.

How credible is this main finding regarding the hypothesis H2-0 on the willingness to help after a performance competition?

The answer to this question is also linked to considerations of statistical power. Table 8 reports the power of this study to detect the effect that relative performance evaluations may have on the average number of answers shared in Part 2 for different standardized effect sizes.[35] In addition to the conventional "large", "medium" and "small" effect sizes that Cohen (1977) first suggested for the social sciences, I calculated standardized effect sizes in two related studies with similar experimental decision contexts[36], using their most relevant outcome variables and treatment comparisons (see Appendix B for a detailed description).

Buser and Dreber (2016) find that average contributions in a Public Goods Game are lower under relative pay concerns (standardized effect size d=0.287, two-sided t-test p=0.001). Carpenter at al. (2010) find that group members provide much less favorable assessments of the quality of their peers' work output after the experience of a competition for relative pay (standardized effect size d=|0.621|[37], two-sided t-test p=0.001). My study has a power of 0.4 to

---

[35] The standardized effect size d is defined in the following way $d = \frac{\mu_{helpbas} - \mu_{helptreat}}{SD_{help}}$. I take this study's sample size N=282 as fixed and determine the power of a two-sided t-test to reject the null hypothesis of no effect at the 0.05 level of statistical significance or higher for different d.

[36] Both studies also have a between-subject design in which behavior under relative *pay* concerns (pay tournament) is compared to behavior in a baseline in which piece-rates are paid. Both studies document that competing for relative pay in a first stage has negative behavioral spillovers to subsequent prosocial or anti-social behavior in groups when the two stages are in no way strategically linked.

[37] Due to the construction of the variable, the stand. effect size is -0.621 (negative 0.621) in Carpenter et al. (2010) but with the same implication: the average assessment of other group members' production quality is much lower under the impression of relative pay concerns in reference groups.

detect a standardized effect size d=0.287 of relative performance concerns on average help and a power close to 1 (0.959) if the true effect size was d=0.62. For what is typically considered a medium effect size of d=0.5 my study would reject the null hypothesis of no effect 85% of the times, with the likelihood of a Type-2 error at 15%.

Table 8 Power of Study for Different Standardized Effect Sizes

source	stand. effect size (d)	power
Cohen (1977)	0.8	0.998
Carpenter et. al. (2012)	0.621	0.959
Cohen (1977)	0.5	0.846
Buser and Dreber (2016)	0.287	0.401
Cohen (1977)	0.2	0.222

Notes. The reported power is the one of a two-sided t-test with a level of significance of at least α=0.05 given this study's sample size of 282 participants.

In other words, the results from this study tell us that it is very unlikely that relative performance concerns—caused by relative performance evaluations or relative performance pay—have a substantial negative effect on the willingness to help others perform better at work.[38]

This experimental study can pin down the effect of relative performance evaluations on group morale because I precisely control one of the major confounding factors that arise in organizational settings outside of the laboratory in which, at least in the medium run, higher relative performances may come with monetary rewards. There are several reasons why the main result is informative about how employees or students outside of the laboratory work together. Experimental reference groups in the sample of students are similar to students' actual "professional" reference groups. Moreover, knowledge-sharing is an important dimension of how co-workers in firms can help the productivity of each other.

In important ways, the experimental decision environment was conducive to uncovering a negative effect of relative performance concerns on the willingness to help others perform better. Firstly, the performance ranking was established on general knowledge and the helping behavior involved sharing knowledge in the same type of general knowledge questions. The spillover from a performance competition should be largest when relative performance rankings are established on a particular job and then co-workers can help each other to perform better on that job. Secondly, helping others in the reference group was costly and there was no monetary benefit to

---

[38] The findings of this study cannot speak as confidently about small effect sizes. Reassuringly, if we think about the policy implications of this work, it would be most important to know about sizable costs that relative performance evaluations may have on how members of a reference group work together.

helping for the person that extends help. This makes helping in my experiment a generous act. Moreover, in the absolute majority of reference groups (95%), the members did not know each other prior to the experiment[39] and the experimental protocol did not reveal anything about the group members' identities other than what is revealed in a portrait. This type of intrinsically motivated help should be most sensitive to changes in the level of competition in reference groups, in particular when strangers are at the receiving end of it. Thirdly, the experimental design ruled out that others could observe individual help behavior. Participants helped privately and made all helping decisions once and at the same time. In this environment, the decision not to help was essentially impossible to detect and so was the decision to help.

Therefore, the fact that I find no evidence that relative performance evaluations negatively affect group morale in this decision environment is particularly informative. There is no reason to believe that relative performance concerns may lower the willingness to help others perform better in many other relevant contexts, e.g. when help behavior is observable by others or when there are monetary rewards to it.

Previous experimental work on on-the-job help typically modeled help as the transfer of money. The task introduced in this experiment measures "real help" behavior of participants who can share knowledge for the benefit of others and the outcome of these help decisions depends on own and others' performances. The knowledge-sharing task is versatile and easy to implement in a laboratory setting. I hope that this task will prove useful for researchers who are interested in studying the determinants of on-the-job help.

There are now several studies, including this one, that come to the conclusion that relative performance ranks make a qualitative difference compared to an environment in which this information is not given but not whether it is provided in public or private (see also Tran and Zeckhauser 2012; Ashraf et al. 2014). A general lesson for feedback design seems to emerge here, namely, that relative performance information, provided in private or in public, will put members of the reference group in a comparative and also competitive mind-set.

This study provides first evidence on how the intensity of competition in reference groups compares under relative performance information and relative pay. There is no evidence that the introduction of relative pay further increases the level of competition in reference groups. This finding suggests that, at least in some contexts, relative performance evaluations may substitute relative pay in tournaments. An interesting avenue for future research is to systematically compare individual effort choices under non-monetary and monetary tournaments and when they interact.

The main result on the willingness to help across experimental conditions advances our understanding of when relative performance evaluations do, and when they do not backfire in

---

[39] Out of 282 participants, 6 participants (2%) answered that they knew one person in his or her reference group and 1 participant (0.3%) answered that he or she knew both members of his or her reference group.

reference groups. Charness, Masclet and Villeval (2014) find that relative performance rankings lead to costly unethical behavior in reference groups because group members want to change an initial ranking to a final ranking that is more favorable for them. The authors interpret this main finding as an expression of group members' competitive preferences and their desire for dominance. The decision environment of this study completely removed the strategic link between knowledge-sharing and the ranking outcome itself and competitive preferences under relative performance evaluations did not have a negative consequence on help behavior. Taken together, these results suggest that the frequency of relative performance evaluations may determine whether ore not they have a negative impact on the way members of reference group work together.

The objectivity of the relative performance assessment may have mitigated a potential adverse effect on the willingness to help the productivity of others. Group members knew on what grounds the ranking was established and that the computer impartially implemented the ranking based on performance at the first general knowledge test. The findings by Breza et al. (2018) lend support to this speculation. They find that the negative effects of relative performance pay in workgroups depend on how transparent it was to co-workers that others in the reference group were more productive. Their results indicate that co-workers who knew that pay differences arose from observable performance differences did not react negatively, either by exerting less effort or cooperating less well. In organizations, relative performance evaluations can also mirror subjective perceptions of superiors. At least employees can perceive this to be the case. It may, therefore, be advisable to transparently communicate criteria and performance metrics on which grounds relative performance evaluations are established. It would be interesting to investigate in future work to what extent subjectivity and transparency of evaluation criteria mediate the effect that relative performance evaluations have on group morale or other workgroup behavior.

My results thus draw attention to the challenge of understanding better when relative concerns among employees do and when they do not backfire in reference groups. I find no evidence that relative performance evaluations, whether or not they entail monetary consequences, may have negative consequences for group productivity. The unambiguous positive lesson from this study is that there is no evidence that relative performance concerns in reference groups lower the intrinsic motivation to help others perform better.

References

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2014. "Awards Unbundled: Evidence from a Natural Field Experiment." *Journal of Economic Behavior & Organization* 100 (April): 44–63. https://doi.org/10.1016/j.jebo.2014.01.001.

Austen-Smith, David, and Roland G. Fryer Jr. 2005. "An Economic Analysis of 'Acting White.'" *The Quarterly Journal of Economics* 120 (2): 551–583.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2013. "Team Incentives: Evidence from a Firm Level Experiment." *Journal of the European Economic Association* 11 (5): 1079–1114. https://doi.org/10.1111/jeea.12028.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–78. https://doi.org/10.1257/aer.96.5.1652.

Besley, Timothy, and Maitreesh Ghatak. 2008. "Status Incentives." *The American Economic Review* 98 (2): 206–211.

Black, Paul W, Andrew H. Newman, Bryan Stikeleather, and Nathan J. Waddoups. 2018. "Performance Feedback Type and Employees Subsequent Willingness to Help Other Employees." *Journal of Management Accounting Research*, October. https://doi.org/10.2308/jmar-52298.

Blanes i Vidal, Jordi, and Mareike Nossol. 2011. "Tournaments Without Prizes: Evidence from Personnel Records." *Management Science* 57 (10): 1721–36. https://doi.org/10.1287/mnsc.1110.1383.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2018. "The Morale Effects of Pay Inequality." *The Quarterly Journal of Economics* 133 (2): 611–63. https://doi.org/10.1093/qje/qjx041.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–3319. https://doi.org/10.1257/aer.20170029.

Buser, Thomas, and Anna Dreber. 2016. "The Flipside of Comparative Payment Schemes." *Management Science* 62 (9): 2626–38. https://doi.org/10.1287/mnsc.2015.2257.

Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm. 2010. "Tournaments and Office Politics: Evidence from a Real Effort Experiment." *American Economic Review* 100 (1): 504–17. https://doi.org/10.1257/aer.100.1.504.

Charness, Gary, Uri Gneezy, and Austin Henderson. 2018. "Experimental Methods: Measuring Effort in Economics Experiments." *Journal of Economic Behavior & Organization* 149 (May): 74–87. https://doi.org/10.1016/j.jebo.2018.02.024.

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The Dark Side of Competition for Status." *Management Science* 60 (1): 38–55. https://doi.org/10.1287/mnsc.2013.1747.

Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas *." *The Quarterly Journal of Economics* 129 (4): 1625–60. https://doi.org/10.1093/qje/qju023.

Cohen, Jacob. 1977. "CHAPTER 2 - The t Test for Means." In *Statistical Power Analysis for the Behavioral Sciences (Revised Edition)*, 19–74. Academic Press. https://doi.org/10.1016/B978-0-12-179060-8.50007-4.

Danilov, Anastasia, Christine Harbring, and Bernd Irlenbusch. 2019. "Helping under a Combination of Team and Tournament Incentives." *Journal of Economic Behavior & Organization* 162 (June): 120–35. https://doi.org/10.1016/j.jebo.2019.03.013.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics* 132 (4): 1593–1640. https://doi.org/10.1093/qje/qjx022.

Drago, Robert, and Gerald T. Garvey. 1998. "Incentives for Helping on the Job: Theory and Evidence." *Journal of Labor Economics* 16 (1): 1–25. https://doi.org/10.1086/209880.

Duffy, John, and Tatiana Kornienko. 2010. "Does Competition Affect Giving?" *Journal of Economic Behavior & Organization* 74 (1–2): 82–103. https://doi.org/10.1016/j.jebo.2010.02.001.

Ellingsen, Tore, and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3): 990–1008. https://doi.org/10.1257/aer.98.3.990.

Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis. 2011. "Relative Earnings and Giving in a Real-Effort Experiment." *The American Economic Review* 101 (7): 3330–48.

Ewers, Mara, and Florian Zimmermann. 2015. "IMAGE AND MISREPORTING: Image and Misreporting." *Journal of the European Economic Association* 13 (2): 363–80. https://doi.org/10.1111/jeea.12128.

Falk, Armin, Anke Becker, Thomas J. Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2725035.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–78. https://doi.org/10.1007/s10683-006-9159-4.

Gächter, Simon, Chris Starmer, and Fabio Tufano. 2015a. "Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale." *PLOS ONE* 10 (6): e0129478. https://doi.org/10.1371/journal.pone.0129478.

Gächter, Simon, Christian Starmer, and Fabio Tufano. 2015b. "Measuring the Impact of Social Relationships: The Value of 'Oneness.'" Mimeo.

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse. 2018. "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision." *Management Science*, February. https://doi.org/10.1287/mnsc.2017.2907.

Karni, Edi. 2009. "A Mechanism for Eliciting Probabilities." *Econometrica* 77 (2): 603–6. https://doi.org/10.3982/ECTA7833.

Köszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707. https://doi.org/10.1162/JEEA.2006.4.4.673.

Kuhnen, Camelia M., and Agnieszka Tymula. 2012. "Feedback, Self-Esteem, and Performance in Organizations." *Management Science* 58 (1): 94–113. https://doi.org/10.1287/mnsc.1110.1379.

Lazear, Edward P. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy* 97 (3): 561–80. https://doi.org/10.1086/261616.

Lazear, Edward P., and Kathryn L. Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." *Journal of Economic Perspectives* 21 (4): 91–114. https://doi.org/10.1257/jep.21.4.91.

Levine, David K. 1998. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1 (3): 593–622. https://doi.org/10.1006/redy.1998.0023.

McManus, T. Clay, and Justin M. Rao. 2015. "Signaling Smarts? Revealed Preferences for Self and Social Perceptions of Intelligence." *Journal of Economic Behavior & Organization* 110 (February): 106–18. https://doi.org/10.1016/j.jebo.2014.12.009.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *The American Economic Review* 83 (5): 1281–1302.

Tran, Anh, and Richard Zeckhauser. 2012. "Rank as an Inherent Incentive: Evidence from a Field Experiment." *Journal of Public Economics* 96 (9–10): 645–50. https://doi.org/10.1016/j.jpubeco.2012.05.004.

Appendix A

Appendix A presents the following additional figures:

1. figures of the feedback screens in the baseline, the private rank feedback condition and the public rank feedback condition
2. pictogram of the Inclusion of the Self in Other (IOS) scale
3. empirical distribution of performance on the timed general knowledge test by experimental condition
4. figures of participants' evaluations of the decision environment in Part 1

Appendix A contains the following additional tables:

1. Results of OLS regressions that show how a participant's confidence in his group members ability to provide correct answers to Part 2 questions changes when relative performance feedback is given. This analysis conditions on the rank of the participant who makes the judgements.
2. List and summary statistics of all questionnaire items from which the indices on attitudes toward cooperation, competition, group work and autonomy are constructed
3. Rank Feedback, Relative Pay and Expected Help, General Cooperativeness and Prosocial Inclinations
4. Results of OLS regressions and Probit models predicting the willingness to share answers with others comparing the baseline to the relative pay condition, controlling for beliefs about correct answers
5. A table that shows the average efficiency loss in groups due to missed opportunities to help and the unrealized gains in group productivity due to missed opportunities to help by experimental conditions

Figure A1. Information Provided at the End of Part 1 (Baseline Condition)
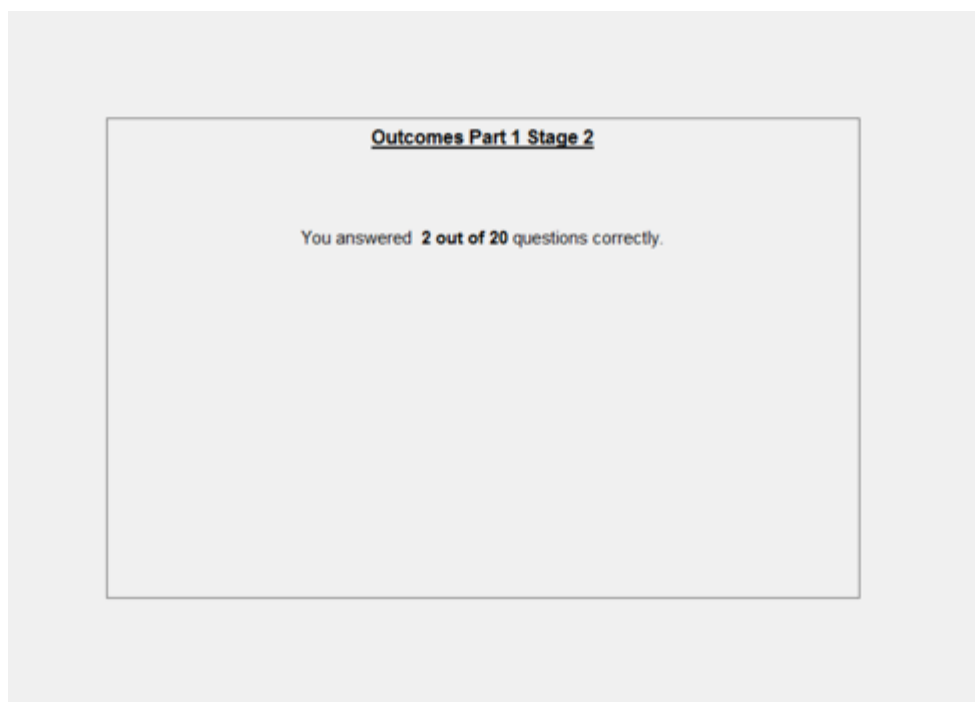
Figure A2. Information Provided at the End of Part 1 (Private Rank Feedback Condition)
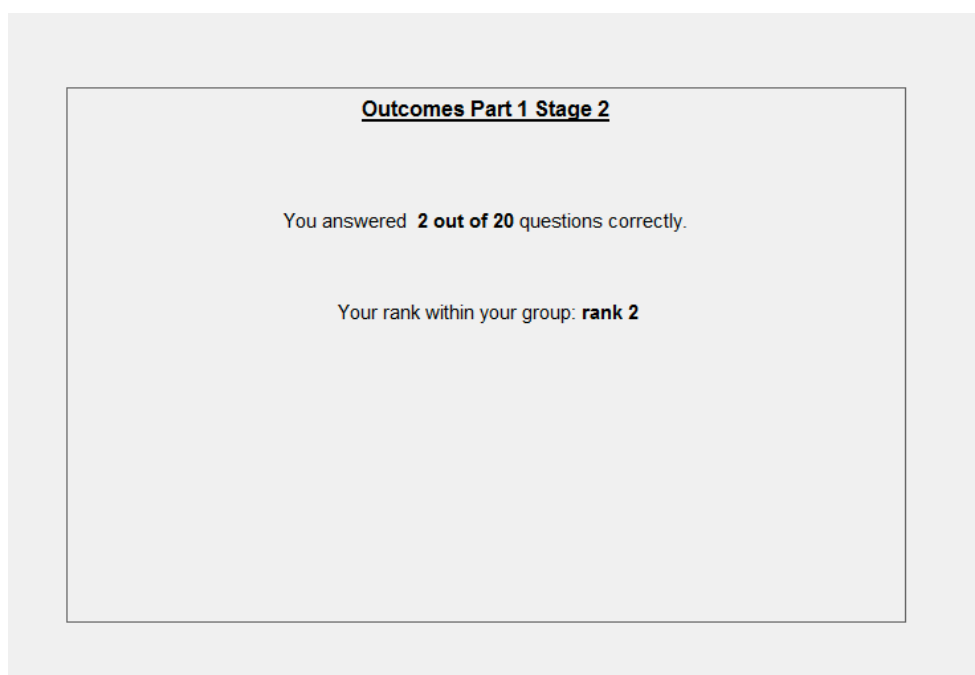
Figure A3. Information Provided at the End of Part 1 (Public Rank Feedback Condition)
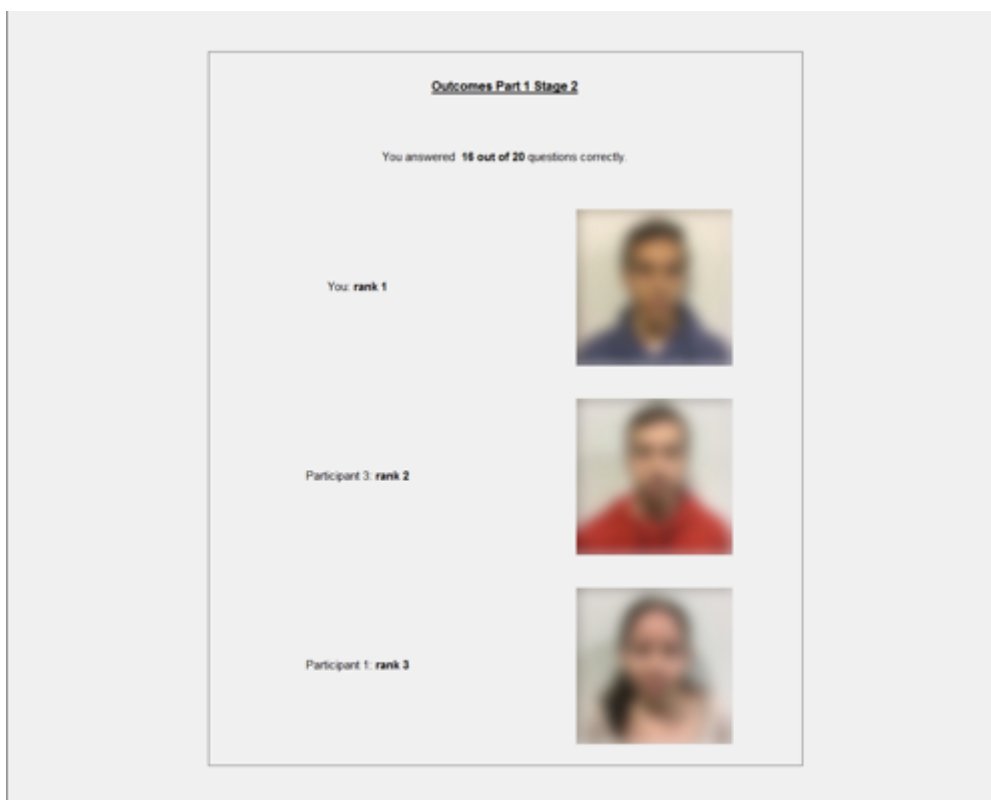
Notes. The faces of participants are blurred here to preserve their anonymity.

Figure A4. Information Provided at the End of Part 1 (Relative Pay Condition)

Figure A5. Help Decision Screen (Public Rank Feedback Condition)

Notes. In all conditions, this screen displayed the portrait of every group member. Across conditions, the help decision screens only varied in the summary of performance on the timed task of Part 1, displayed in the box in the upper right corner of the screen. The screen of the *baseline* condition did only show how many question a participant answered correctly. In the *private rank feedback* condition the box also showed the performance rank of the participant who was looking at that screen. In addition to this information, in *the relative pay* condition, the participant who was looking at the screen also found out whether or not he had obtained an additional bonus payment.

Figure A6. Pictogram of the Inclusion of the Self in Other (IOS) scale
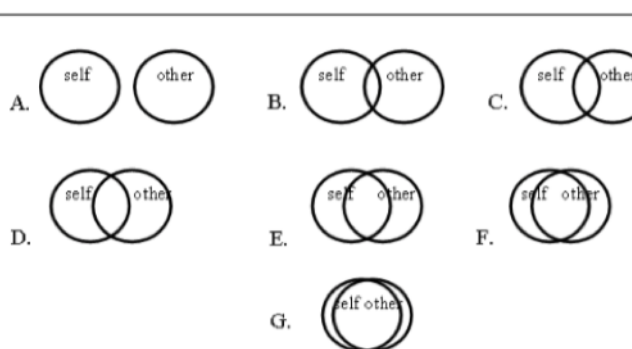
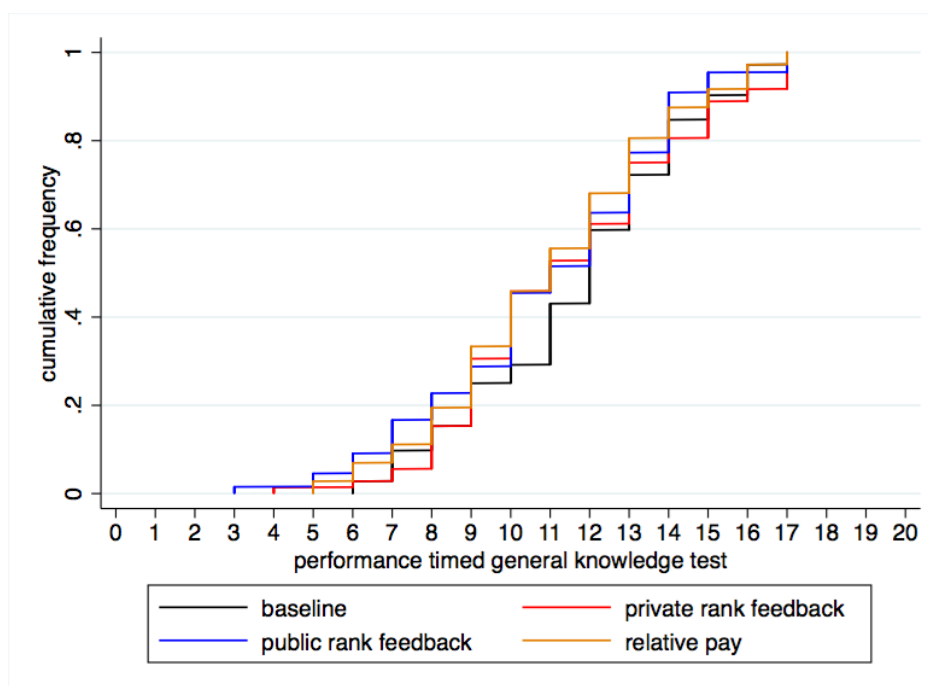Figure A7. Empirical CDF of Performance on Timed General Knowledge Test by Condition

Figure A8. Participants' Evaluations of Part 1 Decision Environment
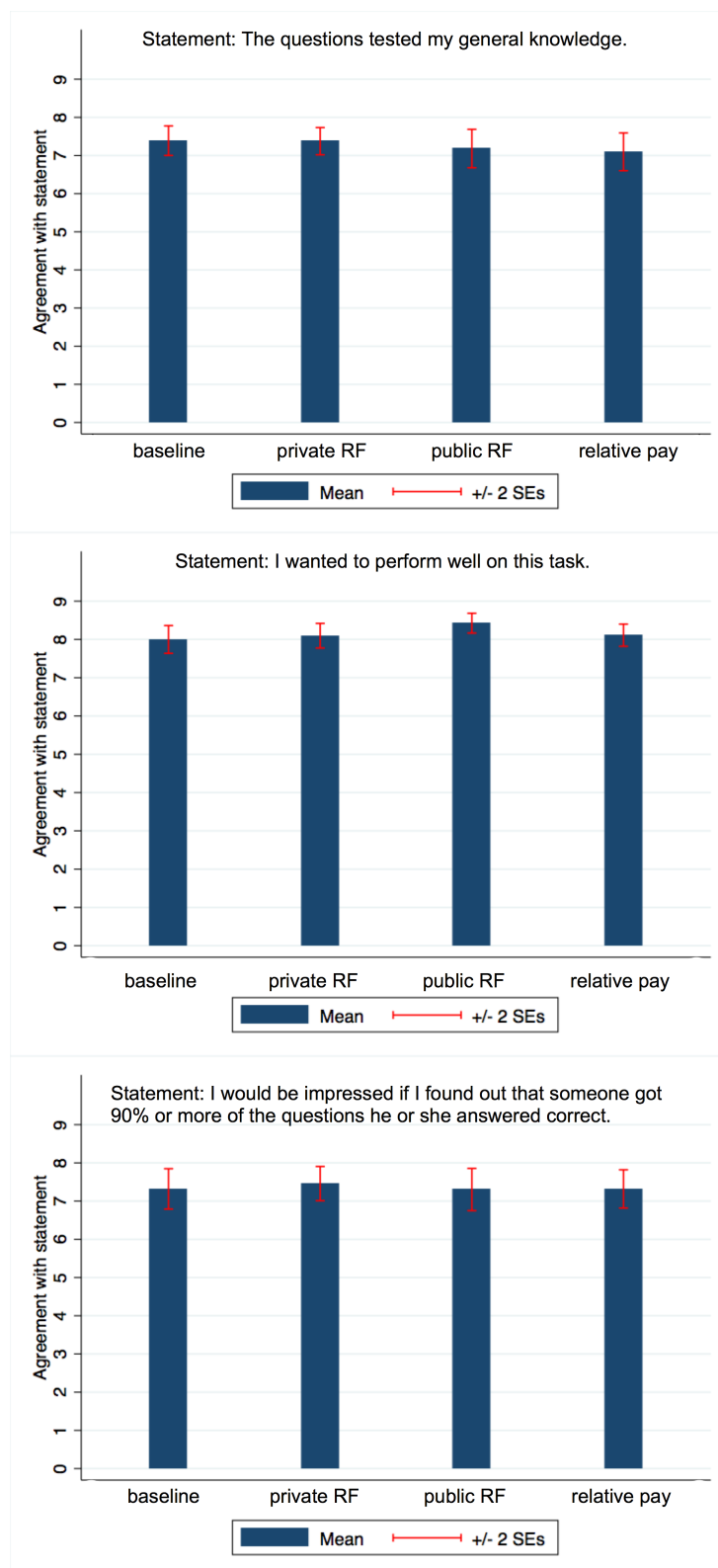
Table A1. Rank Feedback and Confidence in Others' Performance

	OLS predicting belief of participant i of per cent chance that group member (GM) j answered question k correctly						
	Rank 1 participant i			Rank 2 participant i		Rank 3 participant i	
Private RF	-5.546**	-5.262*	Private RF	-0.606	-1.201	10.09****	10.07****
	(2.755)	(2.698)		(3.193)	(3.227)	(2.948)	(2.893)
Public RF	-7.171***	-7.358***	Public RF	2.348	-0.943	11.49****	9.153**
	(2.702)	(2.551)		(3.425)	(3.665)	(3.421)	(3.481)
GM j rank 3		-0.830	GM j rank 1		0.0935		-0.153
		(1.227)			(1.512)		(0.835)
Priv. RF X GM j rank 3		-0.571	Priv. RF X GM j rank 1		1.187		-0.0897
		(1.441)			(1.865)		(1.158)
Pub. RF X GM j rank 3		0.362	Pub. RF X GM j rank 1		6.518****		4.558**
		(2.072)			(1.864)		(1.759)
Controls	*Yes*	*Yes*		*Yes*	*Yes*	*Yes*	*Yes*
Constant	61.44****	61.83****	Constant	25.66***	25.50***	41.81****	41.76****
	(12.04)	(12.04)		(8.827)	(8.788)	(8.193)	(8.261)
Obs.	1400	1400	Obs.	1400	1400	1400	1400
R^2	0.1391	0.1396		0.1585	0.1636	0.1327	0.1354

Notes. Private RF and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *GM j rank 1* and *GM j rank 3* are indicators for whether the group member that participant i judges has performance rank 1 or 3, respectively, on the timed task in Part 1. These indicators are also interacted with the treatment indicators. The following variables are controlled for. *Actual correct* variable indicates whether *participant i* or the *group member j t*hat he judges provided a correct answer to a Part 2 question k. *Performance Part 1* records the number of questions that participant i answered correctly during the timed task in Part 1. *Risk attitude* is where participant i positioned himself on a scale from 0=very risk-averse to 10=very risk seeking. *Female* indicates whether participant i is a woman. *Female group member* indicates whether group member j is a woman. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table A2. Questionnaire Items on Attitudes Toward Cooperation, Competition, Working in Groups and Working Alone

Item	Category	Mean	SD
I am drawn to compete with others.	Competitiveness	4.98	2.07
It annoys me when others perform better than I do.	Competitiveness	5.16	2.15
I feel that winning or losing doesn't matter to me.	Competitiveness (-)	3.86	2.09
I avoid competitive situations.	Competitiveness (-)	4.57	2.24
I love to help others.	Cooperativeness	7.33	1.34
I like to share my ideas and material with others.	Cooperativeness	6.54	1.63
I avoid doing favors to others.	Cooperativeness (-)	2.70	1.73
I expect everyone to look out for themselves.	Cooperativeness (-)	5.57	1.94
I like to work things out on my own.	Autonomy	6.92	1.63
Given the choice, I prefer to work on an assignment alone rather than getting an assignment in which I have to work together with others.	Autonomy	5.37	2.14
I find it hard to work by myself.	Autonomy (-)	2.73	1.55
I find I am less productive when I work by myself.	Autonomy(-)	3.05	1.83
I can learn important things from other colleagues or fellow students.	Groupwork	7.94	1.29
I like working in groups.	Groupwork	5.79	1.97
In workgroups, one person does typically most of the work.	Groupwork (-)	5.88	1.94
I find that working in groups is often inefficient.	Groupwork (-)	5.65	1.97

Notes. This table summarizes all the fours items from which the index for that category is constructed. The answers to each question ranges from 1- does not apply at all to 9- definitely applies. The index is the average score across the four items of a category and negatively keyed items are reverse scored.

Table A3. Rank Feedback, Relative Pay and Expected Help, General Cooperativeness and Prosocial Inclinations

	Expected Help	Cooperativeness	Positive Reciprocity	Negative Reciprocity
Private RF	-0.201	-0.101	-1.008*	-0.771
	(0.389)	(0.175)	(0.522)	(0.744)
Public RF	-0.0789	-0.0571	-0.457	0.00184
	(0.388)	(0.176)	(0.556)	(0.777)
Relative Pay	-0.403	0.0590	-0.485	-0.522
	(0.417)	(0.211)	(0.577)	(0.874)
Constant	4.632****	5.955****	13.63****	10.83****
	(0.215)	(0.124)	(0.391)	(0.513)
Obs.	282	282	282	282
R^2	0.0039	0.0030	0.0093	0.0048

Notes. Private RF and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Expected help* is the number of answers to Part 2 questions that participant i thinks his two group members shared with him on average. Cooperativeness is an index that measures the desire to help others in general and runs from 1 (very low) to 9 (very high). Robust standard errors are in parentheses, 94 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level.

Table A4. Predicting the Willingness to Help Under Relative Pay Compared to Baseline

	Predicting Prob(share answer to question k)			
	Probit		OLS	
Relative Pay	-0.0641	-0.0559	-0.06388	-0.05258
	(0.0478)	(0.0527)	(0.04784)	(0.04883)
Belief correct (self)		0.6442****		0.60298****
		(0.0950)		(0.08113)
Belief correct (others)		-0.2064*		-0.19484*
		(0.1076)		(0.09987)
Constant	Yes	Yes	0.40277****	0.1193**
			(0.0261)	(0.0522)
Obs.	1,440	1,440	1,440	1,440
(pseudo) R^2	0.003	0.088	0.004	0.1109

Notes. Predicted marginal effects at mean level of covariates are reported (Probit). *Share answer to question k* is an indicator for whether participant shared the answer to a Part 2 question k with others. *Relative Pay* is an indicator for participant i privately observing his performance rank on the timed task and the best performer receiving a substantial monetary bonus at the end of the timed task. *Belief correct self* ranges from [0,1] and is the subjective probability that participant i gives to the event that his answer to question k is correct. *Belief correct others* ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to question k. Robust standard errors are in parentheses, 48 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table A5 Wilcoxon-Mann-Whitney Tests

rank	compare number of answers shared across conditions	p-value
1	Baseline-Private RF	0.6474
	Baseline-Public RF	0.7989
2	Baseline-Private RF	0.2118
	Baseline-Public RF	0.2973
3	Baseline-Private RF	0.6070
	Baseline-Public RF	0.728

Table A6. Missed Opportunities to Help, Efficiency and Group Productivity by Experimental Condition

	Average efficiency loss	Average gains in group productivity
Baseline	5.12 CHF	15.2%
Private RF	5.29 CHF	16.3%
Public RF	6.57 CHF	21.7%
Relative Pay	6.27 CHF	20.0%
Total	5.80 CHF	18.2%

Notes. Average efficiency loss due to missed opportunities to help is the amount of unrealized group earnings net the cost of help in a group, averaged over all groups. Average gains in group productivity is the number of missed opportunities in a group over a group's total productivity, averaged over all groups.

Appendix B

Appendix B contains the following additional tables:

1. Results of OLS regressions which show that a participant's confidence in his or her ability to answer questions correctly is not systemtically affected by relative performance feedback

2. Results of Probit estimations that predict the willingness to help, pooling the data from the private and the public rank feedback treatments.

3. Results of a Probit estimation and an OLS regression that predict the willingness to help when beliefs about correct answers are controlled for in a different way than the one presented in the main text

4. Result of an OLS regression in which beliefs about correct answers (self and others) are interacted with the treatment indicators

5. Results of Probit estimations and OLS regressions that predict the probability of sharing an answer to a question separately for each sub-group with the same performance rank, pooling the data from the private rank feedback and the public rank feedback treatments

6. Results of a Probit estimation that predicts the probability of sharing an answer to a question for the subset of participants whose performance at the timed general knowledge task was neither very good nor very bad according to performance percentiles

Appendix B then presents how the benchmark effect sizes from the two related studies used for the power analysis were obtained.

Appendix B ends with screenshots of the instructions and decision screens as they were presented to participants. For this Appendix B, the faces of participants in the pictures are blurred to preserve their anonymity. In the experiment, the faces were not blurred.

Table B1. Rank Feedback and Self-Confidence

	OLS predicting belief of participant i of percent chance that he answered question k correctly		
	Rank 1	Rank 2	Rank 3
Private RF	1.577	-5.330[*]	2.594
	(3.183)	(3.144)	(3.179)
Public RF	-1.295	-0.0875	2.260
	(2.526)	(3.348)	(3.323)
Actual correct	31.53[****]	28.81[****]	30.05[****]
	(2.011)	(2.147)	(2.333)
Performance Part 1	-3.789	-3.090	-0.370[**]
	(2.641)	(2.254)	(0.156)
Risk attitude	0.360	0.295	0.451
	(0.469)	(0.683)	(0.539)
Female	3.736	-1.997	-2.808
	(2.411)	(2.866)	(2.745)
Constant	48.20[****]	52.16[****]	43.98[****]
	(4.171)	(5.421)	(4.444)
Obs.	700	700	700
R^2	0.2542	0.2097	0.2276

Notes. Private RF and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Actual correct* variables indicate whether participant i provided a correct answer to a Part 2 question k. *Performance Part 1* records the number of questions that participant i answered correctly during the timed test in Part 1. *Risk attitude* is where participant i positioned himself on a scale from 1=very risk-averse to 10=very risk seeking. Female indicates whether participant i is a woman. Robust standard errors are in parentheses, group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table B2. Predicting the Willingness to Help—Probit (pooled data)

	Probit Predicting		
	Prob(share answer to question k)		
Relative performance information	-0.019 (0.0404)	-0.0168 (0.0453)	0.00985 (0.0385)
Belief correct (self)		0.6570^{****} (0.0950)	0.7782^{****} (0.13413)
Belief correct (others)		-0.2753^{***} (0.1125)	-0.3657^{***} (0.1427)
Actual correct			-0.0109 (0.0337)
Performance part 1			0.1477 (0.1535)
Expected help			1.4410^{****} (0.1813)
Risk attitude			-0.2463^{****} (0.0761)
Female			-0.0305 (0.050)
Question FE	*No*	*No*	*Yes*
Constant	*Yes*	*Yes*	*Yes*
Obs.	2100	2100	2100
pseudo R^2	0.0003	0.0795	0.3541

Notes. The predicted marginal effects at mean level of covariates are reported. *Share answer to question k* is an indicator for whether participant shared the answer to a Part 2 question k with his group members. *Private RF* and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Belief correct self* ranges from [0,1] and is the subjective probability that participant i gives to the event that the answer he provided to question k is correct. *Belief correct others* ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to question k. Actual correct indicates whether participant i provided the correct answer to question k. *Performance Part 1* records the fraction of questions that participant i answered correctly during the timed task in Part 1. *Expected help* is the fraction of answers to questions k that participant i thinks the two group members shared with him on average. *Risk attitude* is where participant i positioned himself on a scale from 1=very risk-averse to 10=very risk seeking, divided by 10. *Robust* standard errors are in parentheses, 70 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table B3. Predicting the Willingness to Help Controlling for Beliefs About the Value of Help to the Group

	Predicting Prob(share answer to question k)	
	Probit	OLS
Private RF	-0.03821	-0.0366
	(0.0519)	(0.0508)
Public RF	-0.0012	-0.000669
	(0.0522)	(0.0520)
Valuable help	0.2434****	0.246****
	(0.0645)	(0.0636)
Constant		0.306****
		(0.0362)
Obs.	2100	2100
(pseudo) R^2	0.0157	0.021

Notes. For the treatment indicators and control variables, the predicted marginal effects at mean level of covariates are reported (Probit). Share answer for question k is an indicator for whterh participant shared tha nswer to a Part 2 question k with his group members. *Private RF* and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. Valuable help ranges from $[0,2]$. It sums the two conditional probabilities that a participant i assigns to her answer improving the performance of group member 1 and group member 2, conditional on the group member not knowing the answer to that question. Robust standard errors are in parentheses, 70 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table B4. Interacting Treatment Indicators with Beliefs About Correct Answers

Linear Probability Model Predicting Prob(share answer to question k)	
Private RF	-0.00507
	(0.102)
Public RF	0.148
	(0.120)
Belief correct (self)	0.702****
	(0.110)
Belief correct (others)	-0.266*
	(0.150)
Belief correct (self) X Private RF	-0.0573
	(0.185)
Belief correct (others) X Private RF	0.0141
	(0.229)
Belief correct (self) X Public RF	-0.209
	(0.169)
Belief correct (others) X Public RF	-0.00472
	(0.238)
Constant	0.0974
	(0.0704)
Obs.	2100
R^2	0.1055

Notes. Private RF and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Belief correct self* ranges from [0,1] and is the subjective probability that participant i gives to the event that the answer he provided to question k is correct. *Belief correct others* ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to question k. Robust standard errors are in parentheses, 70 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table B5. Predicting the Willingness to Help for each Rank Sub-Group (pooled data)

	Predicting Prob (share answer to question k)					
	Probit			OLS		
	Rank 1	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
Rel. Performance Feedback	-0.0656 (0.0942)	-0.0793 (0.0838)	0.12037 (0.0830)	-0.0602 (0.0891)	-0.0739 (0.0774)	0.109 (0.0736)
Belief correct (self)	0.5703**** (0.1569)	0.6219**** (0.1489)	0.6514**** (0.1197)	0.538**** (0.135)	0.588**** (0.133)	0.616**** (0.0951)
Belief correct (others)	-0.2277 (0.2100)	-0.0878 (0.1813)	-0.4304*** (0.1769)	-0.207 (0.190)	-0.102 (0.170)	-0.418*** (0.154)
Constant	Yes	Yes	Yes	0.280** (0.119)	0.104 (0.0978)	0.102 (0.0722)
Obs.	700	700	700	700	700	700
(pseudo) R^2	0.0518	0.1010	0.0937	0.069	0.128	0.1070
Mean number of answers shared	4.93	3.86	2.91	4.93	3.86	2.91

Notes. The predicted marginal effects at mean level of covariates are reported (Probit). Relative performance information indicates whether participant i obtained any information on relative performances on the timed task. Belief correct self ranges from [0,1] and is the subjective probability that participant i gives to the event that the answer he provided to Part 2 question k is correct. Belief correct others ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to Part 2 question k. Robust standard errors are in parentheses, 70 group clusters allow for correlated observations at the group and at the subject level. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level.

Table B6. Rank Feedback and Willingness to Help for Non-Extreme Performance on Timed General Knowledge Test

	Probit Predicting Prob(share answer to question k)	
	10th performance percentile> timed test performance i <90th performance percentile	25th performance percentile>= timed test performance i <=75th performance percentile
Private RF	-0.0724 (0.0630)	-0.0780 (0.0699)
Public RF	0.02063 (0.0620)	0.0187 (0.0687)
Belief correct (self)	0.6251**** (0.0989)	0.5845**** (0.1056)
Belief correct (others)	-0.18671 (0.1271)	-0.1551 (0.1329)
Constant	Yes	Yes
Obs.	1570	1420
pseudo R^2	0.0877	0.0757

Notes. Predicted marginal effects at mean level of covariates are reported. *Share answer to question k* is an indicator for whether participant shared the answer to a Part 2 question k with others. *Private RF* and *Public RF* are indicators for participant i privately observing his performance rank on the timed task or publicly observing the performance rank of everyone in his group, respectively. *Belief correct self* ranges from [0,1] and is the subjective probability that participant i gives to the event his answer to question k is correct. *Belief correct others* ranges from [0,1] and is the subjective probability that participant i gives to the event that his average group member provided the correct answer to question k. *Significant at the 10% level,** at the 5% level, *** at the 1% level, **** at the 0.1% level

Effect Size Benchmark Calculation: Carpenter, Matthews and Schirm (2010)

The real effort task in this study was to prepare letters for mailing, which included to stuff a letter in an envelope and to write the address on that envelope. The outcome variable that I use in the benchmark effect size calculations is the difference in the assessment of participant i's production quality when it is done by an objective postal worker or by peers of the same reference groups. Whenever this difference is positive, it means that peer group members engage in "quality sabotage". Quality was rated on a scale from [0,1] for one randomly selected envelope.

I calculated the effect size benchmark comparing quality sabotage in the "Tournament" treatment to quality sabotage in the baseline ("Piece-Rate") condition. I chose this treatment because the peer assessments of the quantity and quality of others' output had no effect on the likelihood of winning the relative pay competition, unlike the "Tournament with Sabotage" treatment. The 25 USD bonus was paid to the group member with the highest quality-adjusted output at a real effort task. In this Tournament condition, only the experimenter's quality and quantity assessment of all group members' work output mattered to determine top performer. Also in the piece-rate benchmark the quality assessment of peers had no material consequences.

Effect Size Benchmark Calculation: Buser and Dreber (2014)

The outcome variable that I used for effect size calculations are a participant i's contributions in Public Goods Game (PGG). I selected their treatment that is most similar to my two treatments. In the "Competition with Feedback" treatment, participants compete for relative pay at a real effort task, a slider-task, and observe at the end of this stage whether they won or lost before they make their allocation decisions in the PGG. In the "Competition" treatment, by contrast, participants find out at the very end of the study whether they lost or won the relative pay competition.

In the following, I reproduce screenshots of the instructions and decision screens exactly as they were shown to participants.

Screen 1: Introduction

Introduction for the study

This is a study on decision making. The study has 3 parts.

In addition to a 10 CHF participation fee you will be paid any additional money you accumulate during the study.

The exact amount you get, in addition to the participation fee, will be determined during the study and will depend on your decisions and on the decisions of others. You will be paid in private, in cash, at the end of the study. This means that no other participants will see how much money you earned. We will read the instructions for each part just before you will make your choices for that part. The choices you make in one part will not affect anything that happens in the other parts. This means that you are making independent decisions in each part of the study.

If you have any questions during the study, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the study. Participants intentionally violating the rules may be asked to leave the study without payment.

Screen 2: Instructions for Part 1 – first untimed general knowledge test

Instructions for Part 1 Stage 1

Part 1 has two stages. At the end of the study, the computer will select one of the two stages from Part 1 to count for your payment. The computer will select one stage at random and give equal weight, that is a 50 % chance, to each stage. Your choices in either of the two stages will not affect which one the computer selects.

In the first stage of Part 1 you will be presented with 10 multiple-choice questions to test your general knowledge. For each general knowledge question, the correct answer will be among the 4 answer choices provided. You can select the answer that you think is correct by clicking on the button next to it. You cannot skip questions. If you are unsure, just select the answer that you find most likely to be correct.

You will get 25 points for a correct answer. Points are converted to CHF at an exchange rate, such that 10 points are worth 0.60 CHF and 25 points are worth 1.50 CHF in payment at the end of the study. There is a 50% chance that you will be paid for the questions you are about to answer, if Stage 1 is selected to count for Part 1. The quality of your answers will not affect which of the two stages of Part 1 the computer selects for payment.

Your payoff in points for Part 1 Stage 1 in case it is selected for payment is:

Payoff for Part 1 Stage 1=25*(number of correct questions) points

You will be able to submit your answers to the general knwoledge questions by pressing the button in the bottom right corner of the screen. We will continue with the second stage of Part 1 when everyone has submitted all 10 answers.

Please click "Continue" once the experimenter asks you to do so. If you have a question, please raise your hand and wait for the experimenter.

Continue

Screen 3: Part 1 Stage 1 – First General Knowledge Test Questions 1-5

Part 1 Stage 1- Questions 1-5 out of 10

You can select an answer to a general knowledge question by clicking on the button next to it. If you are unsure, just select the answer that you find most likely to be correct. You will get 25 points for a correct answer if Stage 1 of Part 1 is randomly selected to count for payment.
Press **Submit** when you have answered all questions.

Question 1

Which English law from 1679 already guaranteed protection from arbitrary and unlawful detention at that time?
- ○ Magna Charta
- ○ Habeas Corpus Act
- ○ Petition of Rights
- ○ Bill of Rights

Question 2

In which layer of the Earth's atmosphere is the ozone layer?
- ○ Troposphere
- ○ Stratosphere
- ○ Ionosphere
- ○ Mesosphere

Question 3

Which sub-organisation of the United Nations is in charge of science and culture?
- ○ UNCTAD
- ○ IAEA
- ○ UNESCO
- ○ UNICEF

Question 4

Which group is not a denomination within Islam?
- ○ Sunnis
- ○ Wahhabis
- ○ Copt
- ○ Shia

Question 5

Which of these vitamins is essential for proper bone and teeth development?
- ○ Vitamin C
- ○ Vitamin B
- ○ Vitamn A
- ○ Vitamin D

Submit

Screen 4: Part 1 Stage 1 – First General Knowledge Test Questions 6-10

Part 1 Stage 1- Questions 6-10 out of 10

You can select an answer to a general knowledge question by clicking on the button next to it. If you are unsure, just select the answer that you find most likely to be correct. You will get 25 points for a correct answer if Stage 1 of Part 1 is randomly selected to count for payment.
Press **Submit** when you have answered all questions.

Question 6

In which direction does the Earth turn?
- East
- West
- North
- South

Question 7

In rhetoric, how is the repetition of identical or similar sounds at the beginning of words called?
- internal rhyme
- refrain
- aphorism
- alliteration

Question 8

Which term does not designate a whirlwind?
- Blizzard
- Tornado
- Cyclone
- Typhoon

Question 9

Who painted "The Scream"?
- Gustav Klimt
- Edvard Munch
- James Ensor
- Ferdinand Hodler

Question 10

Which of these languages belongs to the same languag family as Finnish?
- Faroese
- Lithuanian
- Hungarian
- Dutch
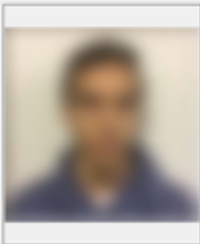
Submit

Screen 5: Confirmation of picture

Screen 6: Introduction of Groups
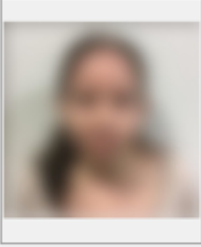
Instructions for Part 1 Stage 2
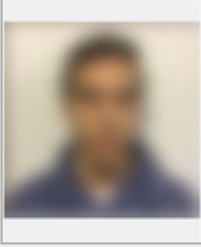
At the onset of today's study you were randomly matched with 2 other participants to form a group. The assignment to groups was entirely random and was not influenced anything that happened in the study so far. You and the two other participants with whom you were matched will stay in a group for the rest of the study. During the study, y may interact with your group members through your computer, but you will not interact directly with any of your group members.
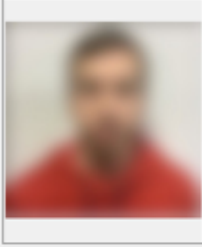
Below, you see the pictures of the three members of your group, including you.

This is your group:

Group Number: 1

Participant 1   Participant 2   Participant 3

Screen 7: Instructions for the Timed task with varying relative performance feedback (by treatment)

Instructions for Part 1 Stage 2 (continued)

In the second stage of Part 1 you will be presented with more general knowledge questions of the same type as the ones you were tested on in the first stage. This time, you will have 3 minutes to answer as many general knowledge questions as you can out of 20. You will see one question at a time and you can proceed to the next question once you have selected and submitted an answer. Note that once you submit the answer to a question you will not be able to return to it and revise it. Everyone sees the same sequence of the same 20 questions. For each general knowledge question, the correct answer will be among the 4 answer choices provided. You cannot skip questions. If you are unsure, just select the answer that you find most likely to be correct.

Payoffs in the second stage of Part 1 are the same as in the first stage of Part 1. If this part is randomly selected to count for payment, you will get 25 points for a correct answer. Points are converted to CHF at an exchange rate, such that 10 points are worth 0.60 CHF and 25 points are worth 1.50 CHF in payment at the end of the study.

Your payoff in points for Part 1 Stage 2 in case it is selected for payment is:

> Payoff for Part 1 Stage 2 =25*(number of correct questions) points

Remember, at the end of the study, the computer will select one of the two stages from Part 1 to count for your payment. The computer will select one stage at random and give equal weight, that is a 50 % chance, to each stage.

Feedback at the end of Part 1 Stage 2

A clock in the upper right corner of your screen will indicate how much time you have left to answer questions. When the 3 minutes have passed you will see a feedback screen. The screen will show:

> A)
> How many questions out of 20 you answered correctly.
>
> B)
> How your answers compare to the answers of the other two participants in your group. You will find out whether you rank first, second or third in your group. This ranking is relative to the correct answers obtained by the others in your group during the timed task. For example, Rank 1 indicates that you answered the most questions correctly in the group. If two group members answered the same number of questions correctly, the computer flips a coin to determine who will be ranked higher.
>
> C)
> You will see the rank of each of the three members of your group next to their pictures. That is, you will know your rank and also exactly where all three members of your group rank relative to each other.

To summarize:
In Part 1 Stage 2 you have 3 minutes to answer as many general knowledge questions as possible out of 20. If this part is randomly selected to count for payment, you will get 25 points for a correct answer. At the end of the 3 minutes you will see a feedback screen showing how many questions out of 20 you got correct and where the number you got correct ranks relative to the correct answers obtained by your group members. You will see the rank obtained by each of the three members of your group, including you, alongside all group members' pictures.
The timed task will begin automatically so that everyone starts at the same time.

Are there any questions?

Notes. These are the instructions for the public rank feedback condition. The instructions for the other two conditions only differed in what was described to the participants they would see in the feedback screen. For example, the instructions of the private rank feedback condition stopped after bullet B) in the textbox in the screen.

Screen 8: Timed task (3 minutes)
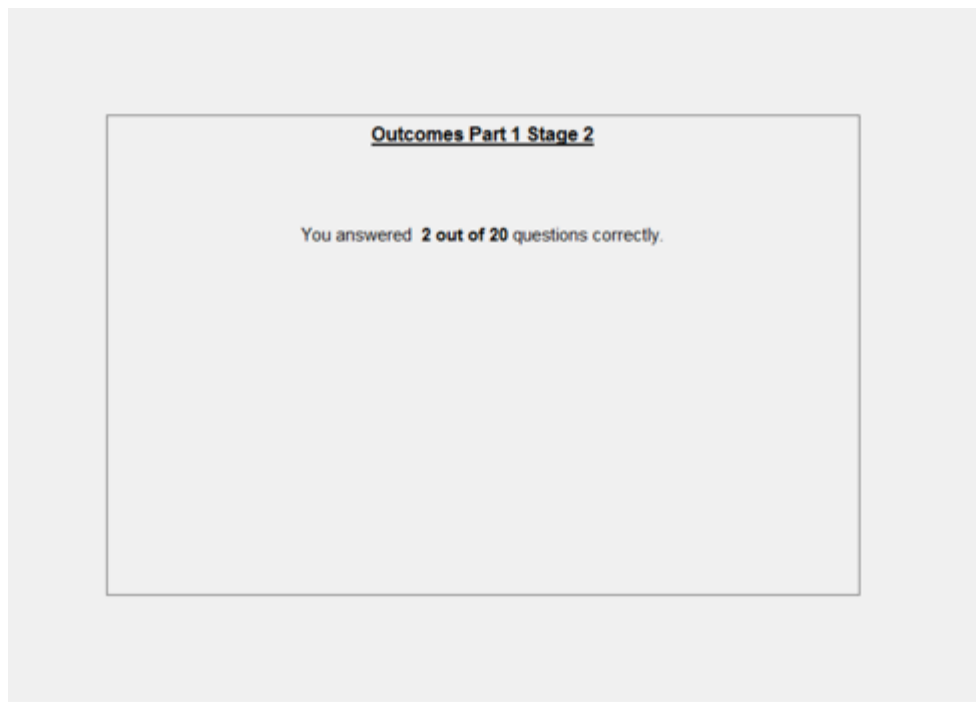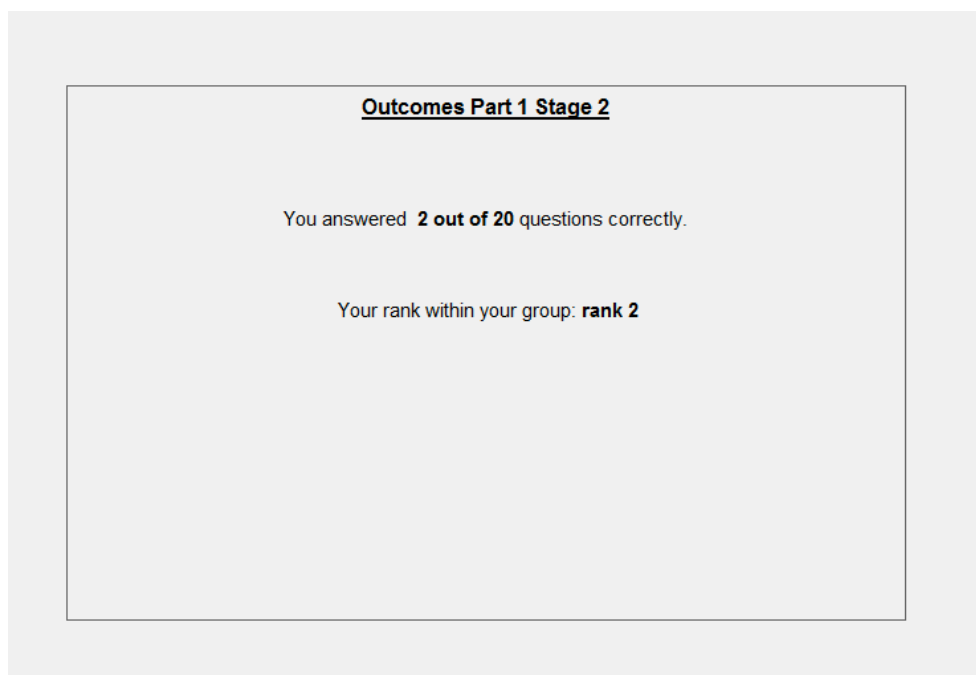

Notes. This is the first question of the timed task. Once a subject submitted an answer to a question the next one appeared on the screen.
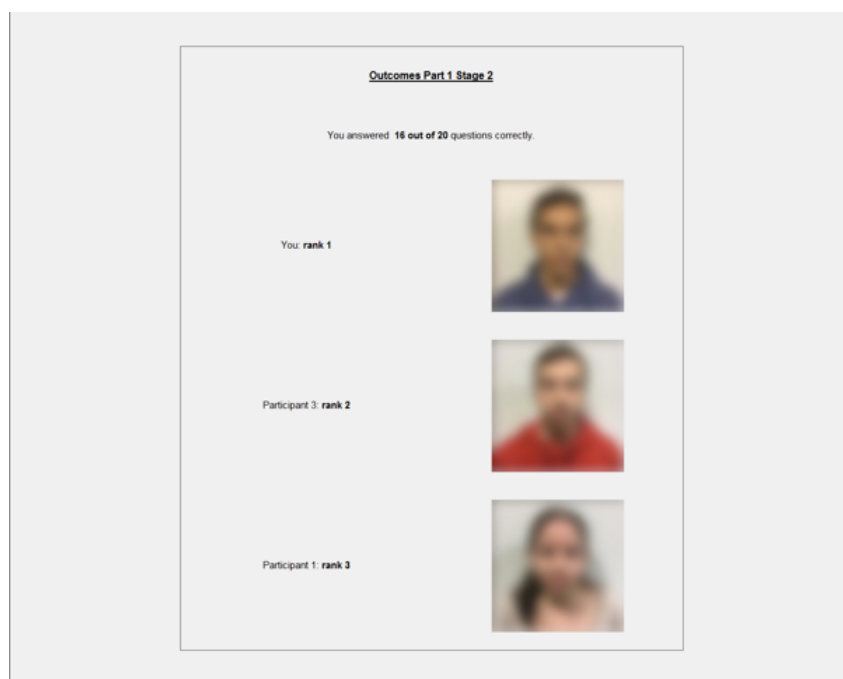
Screen 9: Treatment screen Baseline Condition (feedback about performance at timed task)

Outcomes Part 1 Stage 2

You answered **2 out of 20** questions correctly.

Screen 9: Treatment screen (feedback about performance at timed task) Private Feedback Condition

Outcomes Part 1 Stage 2

You answered **2 out of 20** questions correctly.

Your rank within your group: **rank 2**

Screen 9: Treatment screen (feedback about performance at timed task) Public Rank Feedback Condition

Screen 9: Treatment screen (feedback about performance at timed task) Relative Pay Condition
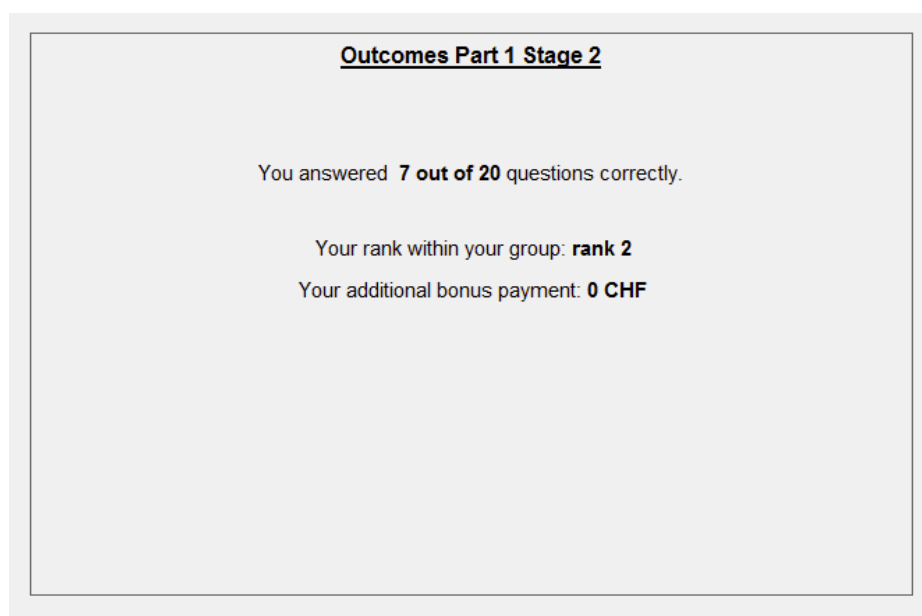
Screen 10: Instructions for the Help Game (Part 2)

Instructions for Part 2

We will now begin with Part 2 of the study. In this part, you and your group members have a chance to earn more money by correctly answering 10 new general knowledge questions of the same type as before. Every group member will get 25 points for each question he or she answers correctly in Part 2. Remember, points are converted to CHF at an exchange rate such that 10 points are worth 0.60 CHF and 25 points are worth 1.50 CHF in payment at the end of the study. When you see the money your earned in Part 2 in your payment overview at the end of the study, you will find out how many questions you got correct. This is the only information you will obtain about choices made in Part 2.

During the task, you can send your answers to your group. This can help your group members to get a question correct. As you complete Part 2, you will select an answer to a question **and** decide whether to send that answer to your group. Here is how this will work. You will see a multiple-choice general knowledge question on the screen in front of you. You can select the answer that you think is correct by clicking on the button next to it. At the same time, you can also decide whether to send your answer to your group members by checking the statement, "Send my answer to question X to my group members.", below where you provide your answer. Sending your answer helps your group members by replacing their incorrect answers with your answer if it is correct.

Example 1 shows how this works.

> Example 1: If a group member, say Group Member 1, provides an incorrect answer then Group Member 1 gets 0 points for that question. However, if either Group Member 2 or Group Member 3 sent a correct answer to the other group members, then that correct answer replaces Group Member 1's incorrect answer. In this case, Group Member 1 obtains 25 points for submitting a correct answer when his or her answers are evaluated for payment.

Sending your answer to your group members will cost you 1 point. This cost does not depend on whether the answer that you send is correct or incorrect. Incorrect answers will not affect your group members in any way.

To summarize:

> • In Part 2 you answer new general knowledge questions and decide whether to send your answers to your group members.
>
> • If you send your correct answer to a general knowledge question it replaces the answer of each other group member who did not get this question correct.
>
> • If you send an incorrect answer it has no effect on your other group members.
>
> • Your payoff in points for Part 2 is:

> Payoff for Part 2=[25*(number of correct questions)-1*(number of answers sent)] points

Your decisions about sending answers to your group do not affect your answers to questions and how they are evaluated for payment. For each general knowledge question, the correct answer will be among the 4 answer choices provided. You cannot skip questions. If you are unsure, just select the answer that you find most likely to be correct.

Here is a quick quiz to make sure you understand how this works. Please raise your hand if you have any questions.

Screen 11: Control Question Help Game

Question 1: How can you get 25 points in Part 2? Select for each answer wheter or not it applies.

By receiving a correct answer to a question from a group member. ○ applies
○ does not apply

By guessing the correct answer to a question. ○ applies
○ does not apply

By sending a correct answer to a question to my group members. ○ applies
○ doesn not apply

By knowing and submitting the correct answer. ○ applies
○ does not apply

Question 2: What happens if you send your answer to a question to your group member who already has a correct answer for that question?

○ This group member's answer only counts as correct if both of us provided correct answers.
○ This group member's answer is replaced with my answer regardless of whether my answer is correct.
○ This group member's answer continues to be correct regardless of whether my answer is correct.
○ My answer is replaced with my group member's answer regardless of whether my group member sent the answer to this question to me.

Continue

Screen 12: Instructions for the Help Game (summary after control questions)

Instructions for Part 2 (continued)
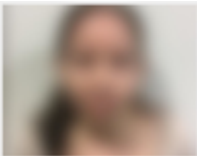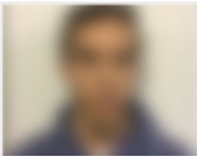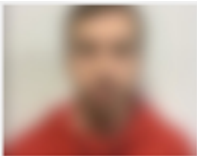
After you have made your decisions about your answers and whether to send your answers to your group, the computer will use your choices to determine which questions you got correct. A question will be considered correct either if you provided a correct answer yourself or if you received a correct answer from one of your group members. You will find out how many questions you got correct in this part, when you see the money you earned in Part 2 in your payment overview at the end of the study. This is the only information you will obtain about choices made in Part 2. You will not find out how many answers were sent to you, who sent answers to you or which questions you got correct.

Your payment for Part 2 will depend on the answers to general knowledge questions that you submitted, the answers that you sent and the answers that your group members sent to you. You will get 25 points for a correct answer, either your own or one that was sent to you. Sending an answer to your group members costs you 1 point.

If you have any questions please raise your hand, otherwise begin now by clicking "Continue".

<div align="right">[Continue]</div>

Screen 13: Help Decisions 1-5

Group 1

Participant 1   Participant 2   Participant 3

Outcomes Part 1 Stage 2

You answered **16 out of 20** questions correctly.

You: **rank 1**

Participant 3: **rank 2**

Participant 1: **rank 3**

You can select an answer to a question by clicking on the button next to it. You can send you answer to your group by marking the checkbox below your answer to that question. Sending an answer costs you 1 point. You will get 25 points for a correct answer, either your own or one that was sent to you. Press **Submit** when you have answered all the questions.

Question 1 of 10

Calling parties "left" and "right" dates back to the seating arrangement in a 19th century parliament. In which country did this parliament convene?

- ○ Germany
- ○ USA
- ○ France
- ○ Russia

□ Send my answer to question 1 to my group members (cost 1 point)

Question 2 of 10

Who wrote the play "Waiting for Godot", which is one of the most important works of the Theatre of the Absurd?

- ○ Oscar Wilde
- ○ Samuel Beckett
- ○ Harold Pinter
- ○ Eugène Ionesco

□ Send my answer to question 2 to my group members (cost 1 point)

Question 3 of 10

Which planet of our solar system is the furthest away from the sun?

- ○ Pluto
- ○ Mercury
- ○ Eris
- ○ Neptune

□ Send my answer to question 3 to my group members (cost 1 point)

Question 4 of 10

Which Greek mythological figure died because he came too close to the sun?

- ○ Daedalus
- ○ Sisyphus
- ○ Icarus
- ○ Tantlaus

□ Send my answer to question 4 to my group members (cost 1 point)

Question 5 of 10
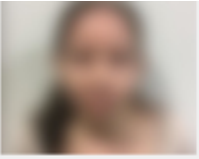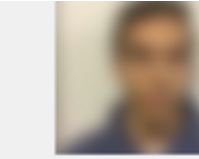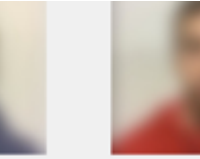
About 200 million years ago, there was only a single supercontinent on earth. What was the name of this supercontinent?

- ○ Pangea
- ○ Gondwana
- ○ Tethys
- ○ Laurasia

□ Send my answer to question 5 to my group members (cost 1 point)

Submit

Screen 14: Help Decisions 6-10

Group 1

Participant 1   Participant 2   Participant 3

You can select an answer to a question by clicking on the button next to it. You can send you answer to your group by marking the checkbox below your answer to that question. Sending an answer costs you 1 point. You will get 25 points for a correct answer, either your own or one that was sent to you. Press **Submit** when you have answered all the questions.

Question 6 of 10

The French Artist Niki de Saint Phalle was known for her colourful, voluptuous female bodies. What did she call her sculptures?
- ○ Naenie
- ○ Nana
- ○ Natal
- ○ Namur

☐ Send my answer to question 6 to my group members (cost 1 point)

Question 7 of 10

The teaching of which pilosopher was the official state religion in China until 1912?
- ○ Mengzi
- ○ Laozi
- ○ Confucius
- ○ Zhuangzi

☐ Send my answer to question 7 to my group members (cost 1 point)

Question 8 of 10

Which UN organisation replaced the GATT (General Agreement on Tariffs and Trade) in 1996?
- ○ OPEC
- ○ OEC
- ○ WTO
- ○ OSCE

☐ Send my answer to question 8 to my group members (cost 1 point)

Question 9 of 10

The subtractive primary colours are colours from which all other colours can be derived in subtractive colour mixing. Which colour does not belong to these primary colours?
- ○ green
- ○ magenta
- ○ yellow
- ○ cyan

☐ Send my answer to question 9 to my group members (cost 1 point)

Question 10 of 10

In linguistics, what is the part of grammar called that deals with the structure of sentences?
- ○ semiotics
- ○ phonemic
- ○ syntax
- ○ semantics

☐ Send my answer to question 10 to my group members (cost 1 point)

Submit

Screen 15: Instructions Belief Elicitation (1)

Instructions for Part 3

We want to know how confident you are in the answers you just gave to the general knowledge questions. We also want to know how confident you are in the answers provided by your group members. The decisions you make here will only affect your own payment. They will not affect the payment of your group members. Nothing you do here can change the answers that were submitted in Part 2 or the earnings from that part of the study.

In this part, you will have the chance to earn additional money by once more submitting answers to the 10 questions from Part 2. You will have the chance to submit 3 answers to each question.

For this part, there are "lotteries" available to help you answer the questions. We have 100 different lotteries and each has a different level of accuracy in providing a correct answer. Each lottery has an accuracy corresponding to an integer between 1 and 100. That is, there is a lottery that provides a correct answer 1% of the time, a lottery that provides a correct answer 2% of the time, a lottery that provides a correct answer 3% of the time, ... , all the way up to a lottery that provides a correct answer 100% of the time. A lottery that provides a correct answer 75% of the time selects the correct answer to the multiple-choice question 75% of the time and a wrong answer to the multiple-choice question 25% of the time.

You have to decide which lotteries you would allow to submit answers for you. For each of the 10 questions from Part 2, you need to make 3 decisions:

      1. Would you rather submit your own answer from Part 2 or have a lottery submit an answer?

      2. Would you rather submit your group member (1)'s answer from Part 2 or have a lottery submit an answer?

      3. Would you rather submit your group member (2)'s answer from Part 2 or have a lottery submit an answer?

Here's how it will work.

You will see your answer to a question from Part 2. Then, you will decide how confident you are in this answer. You will do this by choosing which lotteries you would allow to submit an answer to this question instead of submitting your answer from Part 2. Specifically, you will choose an accuracy threshold (a number between 1 and 100) for your answer such that for any lottery that has an accuracy higher than or equal to your threshold, you would prefer to have the lottery submit an answer instead of submitting your answer from Part 2. This also means that for any lottery that has accuracy lower than your threshold, you would prefer to submit your answer from Part 2 instead of letting the lottery submit an answer. Thus, your accuracy threshold reflects how confident you are in your answer to this question.

You will indicate how confident you are in your answer to a question from Part 2 by filling in the blank with the number that makes the following statement true for you:

"I think that the chance that my answer is correct (in %) is: ___ ."

You will then write in your accuracy threshold for this question. You need to enter a number between 1 and 100.

A lottery will then be randomly selected for that question. Each lottery from 1% to 100% is equally likely to be chosen. If the randomly chosen lottery has an accuracy higher than or equal to your threshold for that question, the lottery will submit an answer to the question for you. If the randomly chosen lottery has an accuracy less than your threshold, your answer from Part 2 will be submitted, instead.

Example 2 shows how this works:

      <u>Example 2</u>: If you choose 75% as your accuracy threshold for a question, and the lottery randomly selected for that question has an accuracy of 90%, this lottery will submit an answer to the question for you. The lottery will have a 90% chance of getting the question correct. If you choose 75% as your accuracy threshold, and, instead, the lottery randomly selected for that question has an accuracy of 20%, your answer from Part 2 will be submitted instead of the lottery's answer.

      Because of this, you are most likely to provide a correct answer when you write exactly what you think the chance is that your answer from Part 2 is correct (i.e. if you believe there's a 75% chance your answer is correct, you should write 75 and not something higher or lower.)

Continue

*Note*s. I modeled these instructions after Coffman (2014) which are publicly available.

Screen 16: Instructions Belief Elicitation (2)

Instructions for Part 3 (continued)

You will also have a chance to submit two more answers for each question. For the second and third answers, you can either submit the answer provided by one of your group members in Part 2, or let a lottery submit an answer. So you will be deciding how confident you are in each of your group members' answers from Part 2. Note that you will not find out the actual answer provided by your group members in Part 2.

You will see a question from Part 2. Then, you will decide how confident you are in your group member's answer to this question. You will do this by choosing which lotteries you would allow to submit an answer to this question instead of submitting your group member's answer. You will choose an accuracy threshold (a number between 1 and 100) for the answer that this person provided such that for any lottery that has an accuracy higher than or equal to your threshold, you would prefer to have the lottery submit an answer instead of submitting your group member's answer from Part 2. This also means that for any lottery that has an accuracy lower than your threshold, you would prefer to submit your group member's answer from Part 2 instead of letting the lottery submit an answer. Thus, your accuracy threshold reflects how confident you are in your group member's answer to this question.

You will indicate how confident you are in your group member's answer to a question from Part 2 by filling in the blank with the number that makes the following statement true for you:

"I think that the chance that Participant X is correct (in %) is: ___ ".

You will then write in an accuracy threshold for this question. You need to enter a number between 1 and 100.

A lottery will then be randomly selected for that question. Each lottery from 1% to 100% is equally likely to be chosen. If the randomly chosen lottery has an accuracy higher than or equal to your threshold for that question, the lottery will submit an answer to the question for you. If the randomly chosen lottery has an accuracy less than the threshold you wrote down for that question, your group member's answer from Part 2 will be submitted instead.

Because of this, you are most likely to provide a correct answer when you write exactly what you think the chance is that your group member's answer from Part 2 is correct (i.e. if you believe there's a 75% chance Participant X's answer is correct, you should write 75 and not something higher or lower.)

Payment

You will submit 3 answers to 10 questions, a total of 30 answers. In this task, you can earn up to 6 CHF that will be added to your payment at the end of the study.

You give 3 types of answers for each of the 10 questions. The first answer is based on how confident you are in your own answer from Part 2. The second answer is based on how confident you are in your group member 1's answer from Part 2 and the third answer is based on how confident you are in your group member 2's answer from Part 2. One answer of each type will be evaluated for payment and you will earn 2 CHF if it is correct. For each type of answer, the computer randomly selects one of the 10 questions. The computer makes independent choices, when it picks one of the 10 questions for each type of answer. For example, the computer could evaluate whether your own or the lottery's answer for question 10 is correct, whether your group members 1's or the lottery's answer for question 2 is correct and whether your group member 2's or the lottery's answer for question 7 is correct. You will earn 2 CHF for a correct answer, regardless of whether the correct answer was provided by you or one of your group members or by the lottery.

Note that you will not know which lotteries have been chosen or what answers are chosen by the lottery. You will also not know the answers that your group members provided in Part 2. Your answers here cannot change your payment from Part 2. The answers you chose in Part 2 will still count when their correctness is evaluated for payment.

Here is a quick quiz to make sure you understand how this works. Please raise your hand if you have any questions.

Screen 17: Control Questions Belief Elicitation

If you write down your accuracy threshold as 61% for a question, and the randomly-drawn lottery is accurate 58% of the time, what will happen?
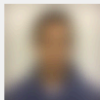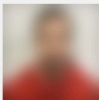
Statement 1: Your answer will be submitted, since the lottery drawn has an accuracy lower than your threshold.

Statement 2: The lottery's answer will be submitted, since the lottery drawn has an accuracy lower than your threshold.

      ○ Statement 1 is correct
      ○ Statement 2 is correct

If you write down your accuracy threshold as 8% for a question, and the randomly-drawn lottery is accurate 10% of the time, what will happen?

Statemnet 1: Your answer will be submitted, since the lottery drawn has an accuracy higher than your threshold.

Statement 2: The lottery's answer will be submitted, since the lottery drawn has an accuracy higher than your threshold.

      ○ Statement 1 is correct
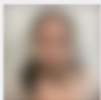   ○ Statement 2 is correct

Continue
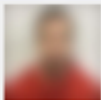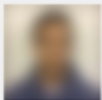
The order in which these beliefs were elicited was fully randomized at the participant level.

Screen 18: Decision Screen Beliefs (1)

In Part 3 we ask you to indicate your confidence in answers to questions from Part 2 by deciding which lotteries you would allow to submit an answer to a question instead of the answer you are evaluating. You can indicate your confidence in an answer to a question provided by you or by each of your group members by filling in the blank with the number that makes the relevant sentence true for you. **You are most likely to provide a correct answer when you write exactly what you think the chance is that the answer you are evaluating is correct.** For each type of answer, the computer randomly and independently selects one of the 10 questions to be evaluated for payment. In Part 3 you will receive 2 CHF for a correct answer, regardless of whether the correct answer was provided by you or one of your group members or by the lottery.

Participant 1   Participant 2   Participant 3

For each type of answer, i.e. an answer provided by you or an answer provided by participant 1 or an answer provided by participant 3, the randomly chosen lottery will submit an answer in case its accuracy is higher than the value that you enter in the box.

Which Greek mythological figure died because he came too close to the sun?
- Daedalus
- Sisyphus
- Icarus
- Tantlaus

I think that the chance that my answer is correct (in %) is:

I think that the chance that participant 1 is correct (in %) is:

I think that the chance that participant 3 is correct (in %) is:

About 200 million years ago, there was only a single supercontinent on earth. What is the name of this supercontinent?
- Pangea
- Gondwana
- Tethys
- Laurasia

I think that the chance that my answer is correct (in %) is:

I think that the chance that participant 1 is correct (in %) is:

I think that the chance that participant 3 is correct (in %) is:

Which is the UN organisation that replaced the GATT (General Agreement on Tariffs and Trade) in 1996?
- OPEC
- OEC
- WTO
- OSCE

I think that the chance that my answer is correct (in %) is:

I think that the chance that participant 1 is correct (in %) is:

I think that the chance that participant 3 is correct (in %) is:

Whose philosopher's teaching was the official state religion in China until 1912?
- Mengzi
- Laozi
- Confucius
- Zhuangzi

I think that the chance that my answer is correct (in %) is:

I think that the chance that participant 1 is correct (in %) is:

I think that the chance that participant 3 is correct (in %) is:

Calling parties "leaft" and "right" dates back to the seating arrangement in a 19th century parliament. In which country did this parliament sit?
- Germany
- USA
- France
- Russia

I think that the chance that my answer is correct (in %) is:

I think that the chance that participant 1 is correct (in %) is:

I think that the chance that participant 3 is correct (in %) is:

Submit 1/2

The order in which these beliefs were elicited was fully randomized at the participant level.

Screen 19: Decision Screen Beliefs (2)

Participant 1 Participant 2 Participant 3

In Part 3 we ask you to indicate your confidence in answers to questions from Part 2 by deciding which lotteries you would allow to submit an answer to a question instead of the answer you are evaluating. You can indicate your confidence in an answer to a question provided by you or by each of your group members by filling in the blank with the number that makes the relevant sentence true for you. **You are most likely to provide a correct answer when you write exactly what you think the chance is that the answer you are evaluating is correct.** For each type of answer, the computer randomly and independently selects one of the 10 questions to be evaluated for payment. In Part 3 you will receive 2 CHF for a correct answer, regardless of whether the correct answer was provided by you or one of your group members or by the lottery.

For each type of answer, i.e. an answer provided by you or an answer provided by participant 1 or an answer provided by participant 3, the randomly chosen lottery will submit an answer in case its accuracy is higher than the value that you enter in the box.

Which planet of our solar system is the furthest away from the sun?
- Pluto
- Mercury
- Eris
- Neptune

I think that the chance that my answer is correct (in %) is: ☐
I think that the chance that participant 1 is correct (in %) is: ☐
I think that the chance that participant 3 is correct (in %) is: ☐

The subtractive primary colours are colours from which all other colours can be derived in subtractive colour mixing. Which colour does not belong to these primary colours?
- green
- magenta
- yellow
- cyan

I think that the chance that my answer is correct (in %) is: ☐
I think that the chance that participant 1 is correct (in %) is: ☐
I think that the chance that participant 3 is correct (in %) is: ☐

How is the part of grammar called in linguistics that deals with the form and structure of sentences?
- semiotics
- phonemic
- syntax
- semantics

I think that the chance that my answer is correct (in %) is: ☐
I think that the chance that participant 1 is correct (in %) is: ☐
I think that the chance that participant 3 is correct (in %) is: ☐

The French Artist Niki de Saint Phalle was known for her colourful, volumptious female bodies. How did she call her sculptures?
- Naenie
- Nana
- Natal
- Namur

I think that the chance that my answer is correct (in %) is: ☐
I think that the chance that participant 1 is correct (in %) is: ☐
I think that the chance that participant 3 is correct (in %) is: ☐

Who wrote the play "Waiting for Godot" one of the most important works in Theatre of the Absurd?
- Oscar Wilde
- Samuel Beckett
- Harold Pinter
- Eugène Ionesco

I think that the chance that my answer is correct (in %) is: ☐
I think that the chance that participant 1 is correct (in %) is: ☐
I think that the chance that participant 3 is correct (in %) is: ☐

Submit 2/2

Screen 20: Expected Help by Others

Participant 1   Participant 2   Participant 3

You will now make two final decisions in today's study in which you will have the chance to make additional money. Afterwards, the study will continue with a questionnaire. Your responses to these questions are important for scientific purposes, so please read them carefully before answering.

In Part 2 of the experiment you and your group members could send answers to questions to one another. You could each send between 0 and 10 answers. We would like you to tell us how many questions, you believe, each of your two group members sent to your group.

If your guess is sufficiently close to the actual number of answers that your group member sent to your group, 1 CHF will be added to your payment at the end of the study. Your estimate is sufficiently close if you guess the actual number of answers sent to your group +/– 1 answer.

For example, if your Group Member 1 sent 6 answers to the group, a guess of 5, 6 or 7 answers sent would be sufficiently close for you to get 1 CHF. All other guesses would be too far off from the truth and you would get no additional money.

I believe the number of answers that participant 1 sent is: _____

I believe the number of answers that participant 3 sent is: _____

Submit

Screen 21: Questionnaire Manipulation Checks and Attitudes

Questionnaire

Next, we would like you to think carefully about the following statements and rate the extent to which each applies to you.

Please rate the following 20 statements from 1 - "does not apply at all", to 9 - "definitely applies".

You answered at least 20 and up to 60 general knowledge questions today.
Statement 1: The questions tested my general knowledge.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 2: I would be impressed if I found out that someone got 90% or more of the questions that he or she answered correct.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

In part 2 of the experiment you were asked to answer as many questions as possible out of 20. You had 3 minutes for the task and received feedback on your performance.
Statement 3: I wanted to perform well on this timed task.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 4: I felt in competition with the other two members in my group when performing this task.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Please continue to rate the extent to which the following statements apply to you.
Statement 5: I like to work things out on my own.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 6: I avoid competitive situations.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 7: I can learn important things from other colleagues or fellow students.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 8: I love to help others.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 9: I am drawn to compete with others.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Statement 10: I like working in groups.

      1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○ definitely applies - 9

Continue to Statements 11-20

Screen 22: Questionnaire Attitudes (continued)

Questionnaire (continued)

Please think carefully about the following statements and rate the extent to which each applies to you.

Please rate the following 10 statements from 1 - "does not apply at all", to 9 - "definitely applies".

Statement 11: It annoys me when others perform better than I do.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 12: I find it hard to work by myself.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 13: I find that working in groups is often inefficient.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 14: I avoid doing favours to others.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 15: I feel that winning or losing doesn't matter to me.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 16: Given the choice, I prefer to work on an assignment alone rather than getting an assignment in which I have to work together with others.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 17: I like to share my ideas and material with others.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 18: I expect everyone to look out for themselves.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 19: In workgroups, one person does typically most of the work.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Statement 20: I find I am less productive when I work by myself.

    1 - does not apply at all   ○ ○ ○ ○ ○ ○ ○ ○ ○   definitely applies - 9

Submit

Screen 23: Positive Reciprocity (Falk et al. 2016)

Questionnaire (continued)

Imagine the following situation: together with a person whom you do not know you won 100 CHF in a lottery. The rules stipulate the following: One of you has to make a proposal about how to divide the 100 CHF between you two. The other one gets to know the proposal and has to decide between two options. He or she can accept the proposal or reject it. If he or she accepts the proposal, the money is divided according to the proposal. If he or she rejects the proposal, both receive nothing.

Suppose that the other person offered the following split: 50 CHF for you and 50 CHF for himself/herself. **Do you accept this split?** ○ I accept
If you do, you will receive 50 CHF and the other person will receive 50 CHF. If you reject, both of you receive 0 CHF ○ I reject

Suppose that the other person offered the following split: 40 CHF for you and 60 CHF for himself/herself. **Do you accept this split?** ○ I accept
If you do, you will receive 40 CHF and the other person will receive 60 CHF. If you reject, both of you receive 0 CHF ○ I reject

Suppose that the other person offered the following split: 30 CHF for you and 70 CHF for himself/herself. **Do you accept this split?** ○ I accept
If you do, you will receive 30 CHF and the other person will receive 70 CHF. If you reject, both of you receive 0 CHF ○ I reject

Suppose that the other person offered the following split: 20 CHF for you and 80 CHF for himself/herself. **Do you accept this split?** ○ I accept
If you do, you will receive 20 CHF and the other person will receive 80 CHF. If you reject, both of you receive 0 CHF ○ I reject

Suppose that the other person offered the following split: 10 CHF for you and 90 CHF for himself/herself. **Do you accept this split?** ○ I accept
If you do, you will receive 10 CHF and the other person will receive 90 CHF. If you reject, both of you receive 0 CHF ○ I reject

Imagine the following situation: you are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. The ride in the stranger's own car takes about 20 minutes and costs the stranger about 25 CHF in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 7 CHF, the most expensive one 35 CHF.

You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give?

○ The bottle for 7 CHF
○ The bottle for 12 CHF
○ The bottle for 17 CHF
○ The bottle for 25 CHF
○ The bottle for 30 CHF
○ The bottle for 35 CHF

Submit

Screen 24: Negative Reciprocity (Falk et al. 2016)

Questionnaire (continued)

Please consider the following situation: You and another person, whom you do not know, both participate in a study where you can decide on how to assign a certain amount of money and thereby determine the outcome. The rules are as follows.

Both participants get an account with 20 CHF. At the beginning, both participants thus own 20 CHF. The other person decides first. She can transfer money to your account. She can transfer any amount: 0, 1, 2 CHF, etc. up to 20 CHF. Each CHF that she transfers to you is tripled by the conductors of the study and booked to your account. After this first stage the other person therefore has 20 CHF minus the amount she transferred to you in her account. You have 20 CHF plus the tripled amount of the transfer of the other person on your account. Now you get to decide: you have the opportunity to transfer money back to the other person. You can transfer any amount up to 80 CHF, depending on how much you have in your account.

This will be the end of the study and the account balances will be final. The other person has in her account 20 CHF minus the amount she transferred to you plus the amount you transferred back. You have 20 CHF plus the tripled amount of what the other person transferred to you minus the amount you transferred back to her. We would like to know how much you would choose to transfer back to the other person, for a given transfer of her to you.

Suppose the other persons transfers 5 CHF to your account. After the first stage yout then own 20+3*5=35 CHF, the other person owns 20-5=15 CHF.
What amount do you choose to transfer back?

Suppose the other persons transfers 10 CHF to your account. After the first stage yout then own 20+3*10=50 CHF, the other person owns 20-10=10 CHF.
What amount do you choose to transfer back?

Suppose the other persons transfers 15 CHF to your account. After the first stage yout then own 20+3*15=65 CHF, the other person owns 20-15=5 CHF.
What amount do you choose to transfer back?

Suppose the other persons transfers 20 CHF to your account. After the first stage yout then own 20+3*20=80 CHF, the other person owns 20-20=0 CHF.
What amount do you choose to transfer back?

How do you see yourself : Are you a person who is generally willing to punish unfair behavior even if this is costly? Please use a scale from 0 to 10, where 0 means you are "not willing at all to incur costs to punish unfair behavior" and a 10 means you are "very willing to incur costs to punish unfair behavior". You can also use the values in-between to indicate where you fall on the scale.

not willing at all to incur costs to punish unfair behavior -0   ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ 10-very willing to incur costs to punish unfair behavior

Submit