Does Relative Performance Information Lower Group Morale?

Lea Heursen*

March 2023

Abstract

Social comparisons are known to increase individual work morale, but because they could foster competition, they may also negatively impact how well co-workers work together. This paper uses a novel experiment to explore this potential tradeoff. The experiment varies whether members of a reference group receive relative performance information on a knowledge task and measures how this affects the willingness to subsequently help the productivity of others by sharing knowledge with them. The findings reveal that relative performance information spurs competition between members of a reference group when compared to the baseline with no such information. Yet, there is no evidence that relative performance information substantially lowers group morale, that is, the willingness to help the productivity of others in the group. These findings advance our understanding of how relative performance concerns among co-workers affect the way they work together.

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

*Universität Zürich, Department of Economics, Current affiliation: Ludwig-Maximilians-Universität München, Department of Economics, Kaulbachstraße 45, 80539 München, lea.heursen@econ.lmu.de, ORCID: 0000-0002-2733-6773
I am grateful to Roberto Weber and Ernst Fehr for their guidance and support in the development of this project. I thank Björn Bartling, Yves Breitmoser, Dirk Engelmann, Yiming Liu, Eva Ranehill, Ernesto Reuben, Karl Schlag, Florian Schneider, Anja Schöttner, Julien Senn, Vanessa Valero, Georg Weizsäcker, Florian Zimmermann and two anonymous referees for their comments. This paper also benefited from discussions with the audience in numerous conference and seminar presentations. Financial support from the Excellence Foundation Zurich, the Zurich Graduate School of Economics through their Director's Grant program and by Deutsche Forschungsgemeinschaft through CRC TRR 190 (project number 280092119) is gratefully acknowledged. All remaining errors are my own.

1. Introduction

Jobs increasingly require that employees work together (Lazear and Shaw 2007; Deming 2017). Therefore, to understand productivity in modern 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. There are many examples of firms, such as Amazon[1] or Yahoo[2], or schools and universities, that provide relative performance information to employees or students. This type of social incentive can increase individuals' productivity (Blanes i Vidal and Nossol 2011; Kuhnen and Tymula 2012; Tran and Zeckhauser 2012; Gill et al. 2018); but how does it affect the way that people work together and could it be detrimental for teamwork and collaboration? An anecdotal example of such potential unintended consequences is when in 2013, Microsoft abolished relative performance evaluations to put "more emphasis on teamwork and collaboration".[3] At Microsoft, the relative performance evaluations were viewed as encouraging competition among employees.

Relative performance evaluations should spur performance competition among co-workers since performance ranks are a rival good, but does this cause them to be less collaborative in other domains? This study offers an empirical answer to this question. First, I investigate empirically whether relative performance information is sufficient to spur competition among members of a reference group. Then, I examine how the relative performance concerns affect the intrinsic motivation to help the productivity of others. This intrinsic motivation is labeled *group morale* in reference to individual work morale which describes the motivation to exert effort to increase own productivity.

Data from a controlled experimental study with a total of 282 participants answer these research questions. In the first part of the experiment, members of a reference group performed a general knowledge test. The treatments varied whether they are exposed to relative performance rankings on that test privately (private rank feedback), publicly (public rank feedback), or not at all (baseline). A supplementary treatment introduced relative pay in addition to (private) relative performance information.

[1] "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

[2] "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

[3] The then Human Resource Chef of Microsoft, Lisa Bruttel, is quoted on this by the newspaper, The Verge, "Microsoft Axes its Controversial Employee-ranking System", https://www.theverge.com/2013/11/12/5094864/microsoft-kills-stack-ranking-internal-structure, accessed April 9, 2018

In the second part of the experiment, group morale was measured in a way that closely resembles the type of help that is important in workgroups: sharing knowledge. Participants engaged in more knowledge work for a fixed individual piece-rate. They could invest in the performance of others in their reference group by anonymously and privately sharing their answers at a small personal cost.[4] Sharing answers could improve others' performance and earnings because the computer automatically replaced their incorrect answers whenever a correct answer was shared.

In the third part of the experiment participants' beliefs about the correctness of their own and their group members' answers to the questions for which help decisions were made were elicited. This was done in an incentivized and incentive-compatible way. Moreover, the sense of competition in reference groups was measured as the extent to which participants agree with the statement that they felt in competition with their group members in the first part of the experiment. Then, a measure of social distance in reference groups was administered. With these data, I investigate motives for knowledge-sharing and evaluate the impact that relative performance information has on social relations in reference groups.

The experimental laboratory offers three major advantages over observational data sets. First, group morale can be precisely measured, unlike in observational settings where on-the-job help is hard to measure and quantify. Second, an important confounding factor found in the example of Microsoft—and other similar settings involving performance comparisons—can be controlled. In these cases, relative performance rankings could, in the medium term, also come with increased promotion opportunities. It is therefore unclear whether competing for high performance ranks, independent of their monetary consequences, can contribute to creating a competitive culture in an organization. Relative performance information is ubiquitous and occurs at a higher frequency than promotions. Therefore, it is important to know whether this feature of how organizations structure performance feedback can, by itself, create a competitive mindset which leads to unintended consequences detrimental to collaboration. Third, the effect that a competitive mindset in and of itself has on the willingness to help the productivity of co-workers can be isolated. A competitive mindset could nourish a taste for being ahead, under which co-workers simply want to maintain advantageous outcome differences in performance. In a review article on dominance and competition, Rustichini (2008, p.653) summarizes the link between competition and relative concerns in the following way: "Humans who participate in a contest with others have strong preferences on relative

[4] Throughout the experiment, nobody observed the help behavior of others. Thus, helping others' productivity is a prosocial act since it comes without monetary or social image benefits for the helper.

outcomes and are ready to translate these preferences into costly choices." A taste for being ahead is hard to disentangle from strategic concerns, under which employees do not invest in the performance of co-workers because it lowers their chance of being promoted. In the experiment, strategic motives are completely shut down by the ranking being established before help decisions are made.

For the design of social and monetary incentives in organizations, it is important to accumulate more evidence on whether preferences over relative outcomes impact choices that are not directly linked to the competition. Related experimental research suggests that organizations may face a tradeoff between incentivizing individuals and cultivating a cooperative environment. Carpenter et al. (2010) find that a competitive payment scheme negatively affects how peers evaluate each other even when their peer feedback has no effect on the chance of winning the pay tournament. Buser and Dreber (2016) find a negative spillover from tournament pay to a completely unrelated prosocial behavior—the contribution in a Public Goods Game. Contrary to that, results reported by Brandts and Riedl (2022) indicate that an extremely competitive market experience has no overall spillover to subsequent play in Public Goods Games and only decreases contributions when the two players have traded in the same market. Relatedly, Brandts, Riedl and van Winden (2009) find that the experience of losing out over a rival subsequently lowers the social value orientation towards a third-party who chose the rival. This study contributes to this research in important ways. First, it isolates whether a non-monetary source of competition—the provision of rank feedback—is sufficient to trigger negative spillovers to other workplace behaviors. Second, on-the-job help is measured with a task in which the outcome of the help decision depends on the knowledge of the person who provides help and of the group member who receives it. This aims to bring the task closer to the psychology of helping others at work. For example, employees may help a colleague's productivity to enhance their self-esteem or confidence.

This study offers several novel insights on the design of relative performance evaluations and productivity in organizations. First, relative performance information suffices to trigger a comparative and competitive mindset with respect to reference groups.

Second, relative performance concerns in and of themselves do not necessarily unravel the willingness to help the performance of others. The data reveal that even after a performance competition, a substantial share of participants help the productivity of others in the group. In all treatment conditions, group members are, on average, as motivated to share knowledge with others at a personal cost as in the baseline where group members did not compete. Participants are much more likely to share answers that they think will improve the performance of their

group members. This is consistent with a measure of group morale. The substantial sample size in each condition makes this statistical inference credible.[5]

A plausible interpretation of the main result is that the objectivity and transparency of the relative performance evaluation may have mitigated a potential adverse effect on group morale. Findings from a related field experiment in an Indian factory lend support to this interpretation (Breza et al. 2018). The negative effects of relative performance pay on individual effort and the ability to cooperate in workgroups depended on how transparent it was to co-workers that others were more productive. Further interpretations of the main finding and its relation to previous results are discussed in the final section of this paper.

This paper explores a potential cost to the provision of relative performance information and consequently contributes to a series of papers that reveal different unintended side effects of relative performance feedback. These unintended side effects include, 1) enabling assortative productivity matching that ultimately lowers average performance in a firm (Bandiera, Barankay and Rasul 2013) and 2) encouraging unethical behavior in the form of cheating or sabotage in a laboratory experiment (Charness, Masclet and Villeval 2014).

This paper also contributes to the empirical literature on what determines on-the-job help, which complements theoretical work on tournaments and help incentives in organizations (Lazear 1989). Early research finds that promotion tournaments and high-powered individual performance incentives both negatively correlate with self-reports of on-the-job help. This finding occurred in banks (Brown and Heywood 2009), in a manufacturing firm (Drago and Garvey 1998), and among groups of physicians (Encinosa et al. 2007). More recent research by Danilov, Harbring, and Irlenbusch (2019) provides causal evidence that manipulating the relative importance of team pay versus relative performance pay in a laboratory experiment changes average helping behavior (measured as a transfer of money in a stated-effort design) in the expected direction.

The rest of the paper is organized as follows. Section 2 outlines the experimental design in detail. Section 3 shows the results which are then discussed in the concluding section 4.

[5] Given the study's sample size per experimental condition, one can calculate the power of a two-sided t-test to reject the null hypothesis that group morale is not affected by a performance competition in reference groups at a level of significance of at least $\alpha = 0.05$ 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 of $d = 0.5$, this study would reject the null hypothesis of no effect 85% of the time with the likelihood of a statistical Type-2 error as low as 15%. See Section 2 of the Online Appendix for a detailed description of these power calculations.

2. Experiment Design

In Part 1 of the experiment, participants took a timed general knowledge test. A between-subject design randomly varied whether participants received only absolute performance feedback or also information on their relative performance rank in a reference group. A complementary treatment introduced relative pay for best performers in Part 1. Part 2 of the experiment measured how these treatments affected the willingness to help the productivity of other group members. The final parts of the experiment, 3 and 4, elicited the sense of competition in reference groups, social cohesion, and several control variables. Table 1 summarizes the timeline of the experiment:

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 about correct answers
Part 4 Questionnaire (measure perception of competition)

2.1. The Real Effort Task

Throughout the experiment, the real effort task was to answer multiple choice 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 and were worth 1.5 CHF[6] in pay at the end of the experiment.

This real effort task was chosen for several reasons: first, participants cared about performing well and being seen as performing well on it—a fact that was established in pre-tests. This is important since two treatments involved the provision of relative performance feedback without any extrinsic rewards for relative performance. Second, the treatments were expected to have little scope to be able to change performance on a specific question. Participants either knew the answer or not. Thus, knowledge should be the primary factor for good performance. If the answer was unknown, the $1.5-pay per correct answer should have motivated participants to exert cognitive effort to eliminate implausible answer choices in the

[6] At the time of the experiment, the exchange rate from CHF to USD was nearly 1 to 1.

baseline condition as well. This way, I attempted to shut down by design an important confound to the main effect under study. In an alternative design in which Part 1 performance and earnings could substantially increase under rank feedback, the willingness to extend prosocial help in Part 2 could be systematically higher than in the baseline simply due to higher earnings. Third, sharing answers to general knowledge questions is a natural and intuitive way to allow for mutual help on this real effort task.

Each general knowledge question included in the study was pre-tested in the same subject pool to ensure that both the composition of questions in terms of difficulty and the field of general knowledge tested was comparable across parts of the study.[7] The final composition was chosen such that participants would get, on average, 60% of the questions right.

2.2. Part 1—Performance and Feedback

Part 1 had two stages. In the first stage, every participant was tested on his general knowledge with 10 multiple choice questions that had to be answered. This measures baseline ability at answering the type of questions that are used throughout the study. For this first ability measure, everything was held constant across experimental conditions. Participants received no information whatsoever on their performance in this first stage.

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. Groups remained fixed for the entire experiment. When groups were introduced, each group member saw every group member's portrait. Portraits were taken at the beginning of a session by an experimenter to ensure that they are very similar in terms of composition.

In this second stage of Part 1, participants had to answer as many general knowledge questions as possible out of 20 within 3 minutes. Questions appeared one at a time and in a fixed sequence. An answer had to be submitted for the next question to appear on the screen. Time pressure was introduced to ensure heterogeneity in performance on this set of questions. It also ensured that exerting effort in the form of answering questions swiftly and not dwelling on unknown answers would pay off. When the three minutes had elapsed, group members were automatically advanced to a feedback screen where the treatment manipulation occurred.

[7] The objective of the pre-tests was to include questions in the main experiment that would be 1) neither too difficult nor too easy, 2) comparable across parts, and 3) 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.

2.3. Experimental Conditions

The feedback screen was on display for one minute and participants could not manually advance. Table 2 summarizes the information shown in each condition and Figure A1 in the Appendix reproduces screenshots.

In the baseline, each group member found out how many of his 20 answers were correct. Group members had no reference point against which to compare this general knowledge score. In the private rank feedback treatment, a group member also discovered how his performance compared to others' performance. The group member found out whether s/he ranked first, second or third in the group. Rankings were based on the number of correct answers, with ties broken at random. This treatment intended to manipulate self-image concerns about general 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 made the relative performance of everyone common knowledge in the group. The pictures were shown again to enhance social image concerns. The public rank feedback treatment made social image in intelligence explicit, while keeping the information about own relative performance the same as under private rank feedback.[8]

The private and public rank feedback treatments 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 (Bursztyn and Jensen 2015; McManus and Rao 2015; Bursztyn, Fujiwara, and Pallais 2017). I considered these findings in the design of the experiment to ensure that performing well and being seen as performing well on the real effort task is desirable for the participants. This was established in a pre-test.

A fourth treatment introduced relative pay in an environment that was otherwise identical to the private rank feedback condition. This is the most parsimonious way to introduce relative pay concerns in this design. The best performing group member received a substantial bonus

[8]With only three group members, the treatment manipulation of public versus private rank feedback is not as stark as it would be if groups were larger. In the public rank feedback, a group member knows the rank of the others with certainty. In the private rank feedback treatment, a group member knows that a group member has a 50% probability to hold a specific rank. This design choice was made to maximize the number of groups per condition, given a budget.

of 5 CHF, in addition to the piece-rate for correct answers. At the feedback stage, a participant observed her/his own performance rank and whether s/he would receive the bonus (see Figure A1 in the Appendix). With this condition, one can test to what extent, *ceteris paribus*, any results change with the domain of relative concerns (i.e., when money is or is not involved).

Table 2. Experimental Conditions

Baseline	Absolute performance feedback after timed general knowledge test	Public Rank Feedback	Baseline + public information about everyone's performance rank in the group
Private Rank Feedback	Baseline + private information about own performance rank in the group	Relative Pay	Private rank feedback + best performer on timed test earns an additional bonus of 5 CHF

2.4. 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 (see Appendix Figure A2 for a screenshot). In Part 2, group morale was measured as group members' willingness to share their answers to 10 new general knowledge questions. The piece-rate for correct answers stayed the same.

For each question, a participant had the option to share her/his answer with the other two group members.[9] This type of task was chosen 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, etc.) can share this information with others to help their productivity.

Whenever a participant shared a correct answer, the computer automatically replaced group members' incorrect answers with the correct one. Sharing wrong answers had no positive or negative effect on group members. This way, I ruled out by design that participants could sabotage others' performance. The main reason is that, in organizations, on-the-job help is typically a task that benefits its recipient. Moreover, this design feature limits the extent to which (under)-confidence in own answers affects helping decisions.

[9] All instructions for participants described the act of sharing answers with the more neutral term "sending answers" to other group members.

The total benefit to sharing an answer was either 0, 25 or 50 points ($0, $1.5 or $3)—depending on whether a correct answer was shared and on how many group members did not get the question right. This benefit went to *others* in the group. Sharing an answer cost a participant 1 point ($0.06). Thus, when a group member shared an answer, s/he was willing to invest 1 point (4% of the piece-rate) in the performance of others.[10] This small cost to knowledge-sharing mirrors the fact that on-the-job help comes with an opportunity cost in organizations. It also ensures that sharing answers with others is, unambiguously, a prosocial act.

Whether someone can help depends on his performance on a question and his group members' performance. Helping can assist a group member to perform better, just like what sharing knowledge with colleagues can achieve in actual teams. Thus, there are no performance-independent numerical benefits to helping, unlike in other tasks that elicit general prosocial behavior, for example, the Public Goods Game. The novel measure aims to directly capture the psychology of helping others at work. For example, helping colleagues may increase self-esteem when it is motivated by the belief of being better at a task than others. Moreover, this measure of group morale does not exclude the fact that relative performance information could have an overall positive effect on team productivity by increasing the efficiency of knowledge-sharing. For example, after observing their high relative performance, a group member may be substantially more willing to help the group. This efficiency-enhancing effect of relative performance information would be excluded by design with a general measure of prosocial behavior (e.g., the dictator game). Having it increases the external validity of this study's potential finding of a tradeoff that managers may face when they provide rank feedback to their employees.

Participants did not obtain direct feedback about others' choices in Part 2.[11] This precludes that group members could seek to enhance their social image in competence or prosociality through knowledge-sharing, which could confound a treatment effect. Moreover, when participants performed in Part 1, they did not know anything specific about later parts. Participants were unaware that Part 2 would entail prosocial choices.

[10] 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 too 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.

[11] 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.

2.5. Part 3—Beliefs

In Part 3 of the experiment, three beliefs for each Part-2-question were elicited: the subjective probability that a participant assigned to his answer and to the answer of each group member being correct. For this, the order of the Part-2 questions was randomized at the subject level. The mechanism to elicit subjective probabilities in an incentive compatible way was described in Karni (2009) and implementation in this experiment closely followed the protocol introduced in Coffman (2014). The process went as follows: there were 100 lotteries available—each equally likely to be chosen—that had an integer-probability on [1,100] of selecting a correct answer to a question, that is, a 1% chance, 2% chance, 3% chance ... up to a 100% chance. For each question, one of these lotteries was randomly selected. 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. 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 his answer to a question is correct with probability X'.

For each question, participants stated three different such cut-off probabilities: one for their own answer and one for each group member's answer. One of each of the three "types" of belief (i.e., self and the two other group members) was randomly selected and evaluated for payment. Participants earned 2 CHF 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.

With these data, I can assess whether participants intended to use help decisions instrumentally to assist their group members. Moreover, these data can be used to account for the pure information effect of rank feedback which may systematically affect the willingness to help.

Participants were also asked to state how much help they expected to have received from each group member. Participants earned 1 CHF when this guess was within a margin of +/- 1 question to how many answers a group member had actually shared.

2.6. Part 4—Questionnaire

In the questionnaire, several measures were collected to assess perceptions of social relations in the reference groups. The intensity of perceived competition was measured using the participant's response 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 ranging from 1-

"does not apply at all" to 9-"definitely applies". "This task" refers to the timed general knowledge test of Part 1. These data were collected about 30 minutes later and serve as a test of whether relative performance feedback fostered a sense of competition. On the same scale, participants also answered questions to assess 1) to what extent performing well on the general knowledge test and being seen performing well by others was desirable and 2) to evaluate whether they thought that the questions tested general knowledge.

In addition, social distance among reference group members was measured with the Oneness index, ranging from "no connection at all" (1) to feeling "at one" with another person (7). This scale is widely used in psychology to measure the closeness of social relationships (Gächter, Starmer, and Tufano 2015). It has been found to predict behavior in economic studies involving decision-making in groups (e.g., Gächter, Starmer, and Tufano 2022). The Oneness index is constructed from responses to the Inclusion of the Self in Other (IOS)-scale and the We-scale, by averaging them. 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 group member (see Appendix Figure A3 for the pictogram). 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 a group member. As a measure of social relations in reference groups, this variable is intended to provide complementary evidence for the change in competitiveness from the baseline to the treatment conditions.

To complement the behavioral measure of group morale used in this study, general attitudes toward cooperation, toward working in groups or working alone, and toward competition were elicited following the procedure described in Duffy and Kornienko (2010). 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" (see Appendix Table A2 for all items and summary statistics). The index is the average of the four responses, with some items reverse-coded. The four statements on attitudes toward competition were taken from Duffy and Kornienko (2010). After this, reported positive and negative reciprocity were elicited with the survey questions described in Falk et al. (2022).
The questionnaire concluded with a few demographic questions and an elicitation of attitudes toward risk and competition. Participants were asked to position themselves on scales from 0 (very risk-averse; not competitive at all) to 10 (very risk-seeking; very competitive).

2.7. Hypotheses About Behavior in the Experiment

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

> Hypothesis 1-0: The sense of competition in the reference group in Part 1 is not affected by rank feedback.

against the alternative hypothesis that

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

The comparison between the private and public rank feedback conditions explores whether the level of competition within the reference group is larger, on average, when social image concerns about general knowledge are strengthened. Ex ante, it is not clear whether social image concerns would substitute or add to the self-image concerns induced by the private rank feedback. How the level of competition and, ultimately, helping behavior differs across these two conditions is, therefore, an empirical question.

A competition in Part 1 of the experiment would activate positional concerns over relative performances, which may lower the participant's motivation to subsequently extend prosocial help to increase the performance of others. The study offers a direct test of the null hypothesis:

> Hypothesis 2-0: The sense of competition under rank 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 rank feedback lowers the willingness to help other group members perform better in Part 2.

The relative pay treatment explores whether any results change moving from a competition for purely image-based rewards to a competition for a monetary reward.[12]

2.8. Experimental Procedures

The experiment was conducted in English at the Laboratory for Experimental and Behavioral Economics at the University of Zurich and programmed in z-Tree (Fischbacher 2007). 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. Treatments were balanced across sessions (three per treatment) and randomly assigned to sessions. Data was collected in September 2017 (baseline, private rank feedback and public rank feedback) and in June 2018 (relative pay). Table 3 lists the number of participants per condition.[13]

An experimenter took portraits of all participants before they took their seat in the laboratory. Participants were called individually by their participant number and were instructed to make a neutral face. The composition was the same for every portrait— only above the shoulders was captured in the image. Participants then 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. This was also approved by the Ethics Committee of the University of Zurich Department of Economics prior to the beginning of the study.

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**

The instructions for the study were displayed on the computer screen in a participant's cubicle (see Section 3 of the Online Appendix for screenshots). An experimenter read the

[12] The relative pay treatment was conducted several months after the other three conditions and under, otherwise, identical conditions for data collection. To the best of my knowledge, there is no prior empirical work on how the intensity of competition compares across the non-monetary and monetary domain. Ex ante, it is not clear whether the image and monetary return to competition complement or substitute one another.

[13] 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.

instructions for a part out loud just before participants made their choices. Before Part 2 and Part 3, participants also answered mandatory comprehension questions.

The computer selected for each participant whether the first or the second stage of Part 1 was selected for payment, giving equal weight to each stage. Earnings in Parts 2 and 3 and the bonus for the top performer under relative pay were always paid out. Average earnings were 40.00 CHF (including a 15 CHF show-up fee).

3. Results

This section begins with results in support of Hypothesis 1-A. Then, I turn to results on the treatment effect of relative performance information on group morale, which leads to the main result of the paper regarding Hypotheses 2-0 and 2-A. The final subsections present findings on behavior in the relative pay condition. All statistical tests are two-sided.

3.1. Do Relative Performance Comparisons Affect Perceptions of Competition?

Rank feedback on the timed general knowledge test mirrored actual performance differences. Only 12% of the groups had any performance ties. Performance varied substantially from 3 to 17 correct answers. The treatment manipulation was successful in making actual performance differences known to group members. There is no evidence that Part 1 performance systematically differs across the four experimental conditions (see Figure A4 in the Appendix, Kruksal-Wallis test $p = 0.4346$).

Participants had to indicate their agreement with the statement “I felt in competition with the other two members in my group when performing the task [the timed general knowledge test]” on a scale from 1=“does not apply at all” to 9=“definitely applies”. In the baseline, the average sense of competition in groups is very low with 2.0 on this 9-point scale (see column 1 of Table 4). It increases markedly under relative performance information: by 3.4 points, on average, under private rank feedback and by 3.8, on average, under public rank feedback (t-tests $p<0.0001$). Whether the performance feedback is public or private appears to not make a sizable difference on the average sense of competition (t-test $p=0.3926$).

Other questionnaire items measured on the same 9-point Likert scale show that the timed task of Part 1 is very relevant for perceptions of competence and that participants valued

performing well.[14] But, as expected, all these other items on the timed task of Part 1 showed no response to the provision of rank feedback.

Relative performance feedback also increased the perceived social distance between group members, as measured with the Oneness index 35 minutes after the end of Part 1. Column 2 of Table 4 presents the results from ordinary least square (OLS) regressions of the Oneness index, averaged over a participant's two group members, on treatment indicators. Compared to the baseline, reported social closeness decreases by an average of 0.51 points after private rank feedback and by an average of 0.40 points after public rank feedback. There is, however, no evidence that *general* attitudes toward working in groups or working alone were systematically affected by the treatments (columns 3 and 4 of Table 4).

These findings indicate that relative performance evaluations changed the perceptions of social relations within the experimental reference groups. They can be summarized in the first result of the paper.

Result 1: The results support hypothesis H1-A: rank feedback causes perceptions of competition between group members.

Table 4. Perception of Competition, Social Closeness and General Attitudes on Groupwork after Rank Feedback or Relative Pay

	Perception of competition in group	Social closeness with group members	General attitude towards working in groups	General attitude towards working alone
Private RF	3.403***	-0.514***	-0.243	0.191
	(0.337)	(0.181)	(0.229)	(0.187)
Public RF	3.792***	-0.399*	0.155	0.072
	(0.364)	(0.206)	(0.215)	(0.200)
Relative Pay	4.014 ***	-0.469**	-0.215	0.184
	(0.384)	(0.205)	(0.234)	(0.230)
Baseline	2.042 ***	3.028***	5.083***	6.038***
	(0.153)	(0.130)	(0.154)	(0.122)
Obs.	282	282	282	282

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. *Relative Pay* indicates that participant i was in the treatment that paid a bonus to the best performer in addition to providing rank feedback in private. Perception of competition ranges from 1 (no sense of competition at all) to 9 (very strong sense of competition). 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

[14] Participants generally agreed that 1) the general knowledge questions did, in fact, measure their general knowledge (mean agreement = 7.26, SD = 1.83), 2) they wanted to perform well in the Part 1 performance stage (mean agreement = 8.15, SD = 1.31) and 3) they 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).

positive attitude). Robust standard errors are in parentheses, 94 group clusters allow for correlated observations at the group and subject level. *Significant at the 10% level, ** at the 5% level, *** at the 1% level.

3.2. Do Increased Perceptions of Competition Affect Group Morale?

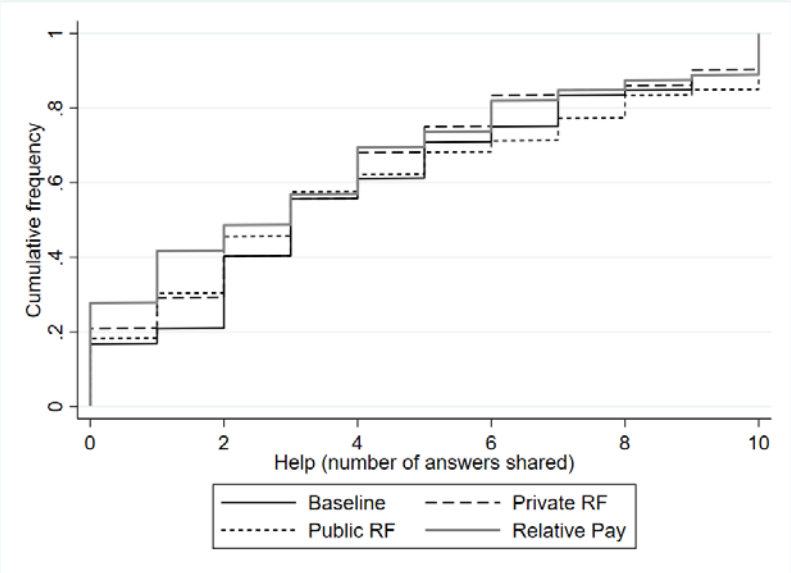
In the following section, I present results from analyzing help behavior in Part 2 of the experiment. In the baseline condition, participants shared on average 4.03 answers out of 10. The estimated average treatment effect of private rank feedback is a decrease in 0.36 answers shared ($-0.11\ \sigma_{help}$) with a 95% confidence interval of [-1.41 answers, 0.69 answers] ([-0.43 σ_{help}, 0.21 σ_{help}]) and p=0.50 (t-test). Effect sizes in terms of standard deviations of helping behavior are reported in parentheses. The estimated average treatment effect of public rank feedback is even smaller, a decrease in 0.013 questions shared ($0.00\ \sigma_{help}$) with a 95%-confidence interval of [-1.15 answers, 1.12 answers] ([-0.34 σ_{help}, 0.34 σ_{help}]) and p=0.98 (t-test).

The empirical distributions of the number of answers shared look very similar across the baseline and the private and public rank feedback conditions (see Figure 1). Wilcoxon-Mann-Whitney tests fail to reject the null hypothesis that these samples of help behavior are drawn from the same population (p>0.53). This provides further evidence that the provision of relative performance information did not systematically change motives for knowledge-sharing.

Consistent with the results on behavior, there is no evidence that the provision of relative performance information caused a shift in the average belief about others' help behavior, compared to the baseline (mean=4.03 answers, SD=3.22).[15]

[15] Difference in means tests have p>0.6.

Figure 1 Empirical CDF of Help by Experimental Condition



In a linear probability model that predicts the willingness to share the answer to a question *k* as a function of treatment status, one can control for beliefs about own and others' correct answers to question *k*. In principle, relative performance information could also change these beliefs, thereby counteracting a negative treatment effect.[16] The first specification in Table 5 predicts the willingness to help as a function of treatment status only. It confirms the results from parametric and non-parametric hypothesis testing: the estimated coefficients of the rank feedback indicators are -0.0361 (private rank feedback) with a 95%-confidence interval of [-0.135, 0.062] and -0.0013 (public rank feedback) with a 95%-confidence interval of [-0.106, 0.103]. The second specification adds controls for beliefs about correct answers: the probability that a participant assigns to his own answer to question k being correct (*belief correct (self)*) and to the answer of his average group member *(belief correct (others)*). The willingness to share the answer to a question increases by an estimated 59 ppt, on average, moving from *belief correct (self)* of 0 to 1, conditional on treatment status and *belief correct (others)*. The coefficient estimate is -23 ppt for *belief correct (others)*. The signs of the two coefficients indicate that participants are more willing to help when they believe that this is valuable. Importantly, the inclusion of these control variables does not change the inference on the two treatment indicators, which are still estimated to not affect the willingness to share answers with others, on average. Note that this result appears not to be driven by the relatively small

[16] Average confidence in own ability to answer questions correctly is not systematically affected by the knowledge of own performance rank, see results presented in Table B1 in the Online Appendix.

cost of sharing since even very confident participants share their answers less than two thirds of the time (see Section 1 and Table B2 of the Online Appendix for this robustness check).

Heterogeneity in Treatment Effect by Rank in Competition

Looking at average treatment effects could mask substantial heterogeneity in how strongly group members with different ranks change their help behavior after rank feedback. Appendix Figure A5 displays average help by rank and by condition. For rank 1 and rank 2 group members, average help is slightly lower after rank feedback compared to the baseline. It is slightly higher for rank 3 group members. However, Wilcoxon-Mann-Whitney tests that compare the distribution of help in the baseline and each treatment conditional on rank provide no indication of a significant heterogeneity in treatment effects (see Appendix Table A1 for p-values). This inference is confirmed with estimates of linear probability models in which the willingness to help conditional on rank is predicted as a function of treatment status and beliefs about correct answers (see Table B3 in the Online Appendix).

This leads to the main result of the paper:

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

To summarize, while the provision of relative performance feedback reliably changes perceptions of social relations in reference groups, that is, the sense of competition and reported social distance between group members, there is no evidence of a systematic change in help behavior.

Relative Pay

The relative pay condition was implemented after observing Main Result 2 to explore whether the domain of competition (monetary versus non-monetary) may matter for relative outcome concerns to lower the willingness to help. This extension was motivated by related studies which found that a pay tournament entails a negative spillover to workplace behavior. Several months passed between rounds of data collection. All other aspects (e.g., subject pool, laboratory etc.) were held constant across rounds. Yet, the results comparing behavior under relative pay to behavior in the other three conditions must be interpreted cautiously. They are only indicative of the effect that the introduction of relative pay has in this decision environment since an additional effect of month-of-data-collection cannot be ruled out.

Under relative pay, the average sense of competition in the reference groups is 6.1 on the 9-point scale, which is comparable to the one under private or public rank feedback (t-test p>0.17). Reported social closeness in the reference groups is, on average, 0.47 points lower than in the baseline (see column 2 of Table 4). Help behavior in the relative pay condition looks very similar to the three other ones (see Figure 1). There is no evidence that competing for relative pay has a sizable negative (or positive[17]) effect on the average willingness to help (see Table 5 for estimates from linear probability models). The overall takeaway is that perceptions and behavior within reference groups appear to be similar during and after a competition for rank and one for relative pay.

Table 5. Linear Probability Models Predicting the Willingness to Help

Private RF	-0.0361	-0.0349
	(0.0496)	(0.0513)
	[-0.135, 0.062]	[-0.137, 0.067]
Public RF	-0.0013	0.0046
	(0.0526)	(0.0536)
	[-0.106, 0.103]	[-0.102, 0.111]
Relative Pay	-0.0639	-0.0529
	(0.0476)	(0.0483)
	[-0.159, 0.031]	[-0.149, 0.043]
Belief correct (self)		0.5894****
		(0.0650)
Belief correct (others)		-0.2280***
		(0.0816)
Constant	0.4028****	0.1492****
	(0.0261)	(0.0427)
Obs.	2800	2800
R^2	0.003	0.100

Notes. The outcome variable *share answer to question k* is an indicator for whether the participant shared the answer to 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. *Relative Pay* indicates that participant *i* was in the treatment that paid a bonus to the best performer in addition to providing rank feedback in private. *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, 94 group clusters allow for correlated observations at the group and subject level. 95%-confidence intervals in square brackets. *Significant at the 10% level, ** at the 5% level, *** at the 1% level, **** at the 0.1% level.

[17] 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 when compared to the best performers in the baseline condition (Wilcoxon-Mann-Whitney test p= 0.9629)..

4. Discussion & Concluding Remarks

This study was designed to empirically test for a detrimental spillover effect of relative performance concerns on group morale in reference groups. The experiment varied by treatment whether individuals were exposed to relative performance rankings of the kinds used in many organizational and social environments. These rankings were based on performance in a general knowledge test. Participants in a group then later made prosocial decisions involving similar general knowledge questions. A group member decided whether to help others' productivity and earnings by sharing his or her answers with them at a small cost. I find that relative performance information causes a large and lasting increase in the sense of competition in reference groups when compared to the baseline condition with no such information. Despite this, there is no indication that relative performance concerns from competition sizably lower the willingness to help others in the group. Beliefs about the value of own help for others is a strong predictor of actual help behavior. This is consistent with the interpretation that participants share knowledge to increase their group members' performance.

How credible is the main finding regarding H2-0 on the willingness to help after a performance competition? The answer is also linked to considerations of statistical power. To guide the power analysis, I calculated benchmark standardized effect sizes[18] in two related studies with similar experimental decision contexts (see Section 2 of Online Appendix for a detailed description of the benchmarking and power analysis). After a competition for relative pay, group members provide much less favorable assessments of the quality of their peers' work output—standardized effect size d=-0.586 (Carpenter et al. 2010). Subjects also lower their average contributions in an unrelated Public Goods Game—standardized effect size d=0.288 (Buser and Dreber 2016). Given this study´s sample size, a two-sided t-test has a power of close to 1 (0.94) to detect a standardized effect size of d=|0.586| at the 5% significance level. If the true effect size was d=0.288 this test would have a power of 0.40. For what is typically considered a medium effect size in the social sciences (d=0.5) (Cohen 1977), the power of this t-test is 0.85, such that the likelihood of a statistical Type-2 error is only 15%. In conclusion, the power considerations, given the sample size, and the results from this study tell us that it is very unlikely that the relative performance concerns associated with relative

[18]The standardized difference in means between two groups is used as an effect size measure, also known as Cohen´s d.

performance feedback have a substantial negative effect on the willingness to help the productivity of others at work.[19]

There is no evidence that performance on the timed general knowledge test improved in the face of relative performance information or relative pay when compared to the baseline. This only seemingly contrasts with the finding in the literature that rank feedback can motivate individual effort and, hence, increase performance. The fact that rank feedback did not increase performance was expected and is a feature of the design (see also section 2.2). First, knowledge is the main factor for good performance. Second, even in the baseline condition, participants faced a high-powered performance incentive ($1.5 piece-rate) that should have encouraged them to swiftly answer questions. The treatments had, therefore, only a very limited scope to be able to further increase performance. This way, average experimental earnings from Part 1 were successfully held constant across conditions, eliminating a potential confound.

In important ways, the experimental decision environment was conducive to uncovering a negative effect of relative performance concerns on the willingness to help others' productivity. First, helping others was a generous act. Intrinsically motivated help should be particularly sensitive to changes in the perceived level of competition in groups. Second, it was virtually impossible to detect others' helping behavior. Thus, several motives for on-the-job help were ruled out by design and for internal validity (e.g., direct reciprocity and the desire to signal prosociality or competence to others). Lastly, experimental reference groups in the student sample are similar to students' actual "professional" reference groups. These design features make the main result particularly informative. For many other relevant contexts, there is no reason to believe that organizations face a tradeoff between relative performance evaluations and group morale. Examples are situations with monetary rewards to helping colleagues' productivity or when help behavior is observable.

Experimental work on prosocial workplace behavior (e.g., on cooperation or on-the-job helping) often models the prosocial behavior as a transfer of money. The task introduced in this experiment aims to bring behavior closer to the psychology of helping colleagues outside the laboratory where, for example, the desire to prove competence may motivate it. Moreover, knowledge-sharing is an important dimension of how co-workers in firms and outside the laboratory can help each other's productivity. This task can prove useful for researchers who

[19] 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.

are interested in studying factors that determine or encourage people's willingness to share their knowledge.

There are now several studies that conclude that relative performance rankings make a qualitative difference when compared to an environment without them but that there is little evidence that public rankings matter more than private ones (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 competitive mindset.

This study's main result advances our understanding of when relative performance evaluations do and do not backfire in 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 more favorable final one. The decision environment of this study completely removed the strategic link between knowledge-sharing and the ranking outcome to test whether activating competitive preferences can in and of itself backfire. Taken together, these results suggest that the frequency of relative performance evaluations may determine whether they have a negative impact on the way members of a reference group work together.

The objectivity of the relative performance assessment, implemented by a computer, may have mitigated a potential adverse effect on the willingness to help others. The findings by Breza et al. (2018) lend support to this speculation. They find that unintended consequences of relative performance pay in workgroups depend on how transparent it was that co-workers were more productive. In organizations, relative performance evaluations can be (perceived as) relatively subjective. It may, therefore, be advisable to make evaluation criteria and performance metrics known. In future work, it would be interesting to investigate to what extent the subjectivity of evaluation criteria mediates the effect that rank feedback has on group morale (or other workgroup behaviors).

This study's 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. The unambiguously positive lesson from this study is that there is no evidence that relative performance concerns in reference groups substantially 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

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. https://doi.org/10.1257/aer.98.2.206

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

Brandts, Jordi, and Arno Riedl. 2022. “Market interaction and efficient cooperation.” *European Economic Review*. 121 (January). https://doi.org/10.1016/j.euroecorev.2019.103318

Brandts, Jordi, Arno Riedl, and Frans van Winden. 2009. “Competitive Rivalry, Social Disposition, and Subjective Well-being: An Experiment”. *Journal of Public Economics* 93: 1158-1167. https://doi.org/10.1016/j.jpubeco.2009.07.010

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

Brown, Michelle, and John Heywood. 2009. "Helpless in Finance: The Cost of Helping Effort Among Bank Employees." *Journal of Labor Research* 30: 176-195. https://doi.org/10.1007/s12122-009-9063-8

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.

Bursztyn, Leonardo, and Robert Jensen. 2015. “How Does Peer Pressure Affect Educational Investments?” *The Quarterly Journal of Economics* 130 (3): 1329–67. https://doi.org/10.1093/qje/qjv021

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, 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

Encinosa William, Martin Gaynor, and James Rebitzer. 2007. "The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems." *Journal of Economic Behavior & Organization* 62:187–214. https://doi.org/10.1016/j.jebo.2006.01.001

Ewers, Mara, and Florian Zimmermann. 2015. “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 Dohmen, David Huffman, and Uwe Sunde. 2022. “The Preference Survey Module: A Validated Instrument for Measuring Risk, Rime, and Social Preferences.” *Management Science* 0(0). https://doi.org/10.1287/mnsc.2022.4455

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. 2015. “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, Chris Starmer, and Fabio Tufano. 2022. “Measuring "Group Cohesion" to Reveal the Power of Social Relationships in Team Production”. *IZA Discussion Papers No. 15512*. http://hdl.handle.net/10419/265733

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* 65 (2): 494-507. 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

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

Rustichini, Aldo. 2008. “Dominance and Competition.” *Journal of the European Economic Association* 6 (2–3): 647–656. https://doi.org/10.1162/JEEA.2008.6.2-3.647

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

The Appendix presents the following additional figures:

1. Figure A1 reproduces the feedback screens in the baseline, the private rank feedback, public rank feedback and relative pay conditions
2. Figure A2 reproduces the help decision screen
3. Figure A3: Pictogram of the Inclusion of the Self in Other (IOS) scale
4. Figure A4: Empirical distribution of performance on the timed general knowledge test by experimental condition
5. Figure A5: Average help by treatment and rank in competition

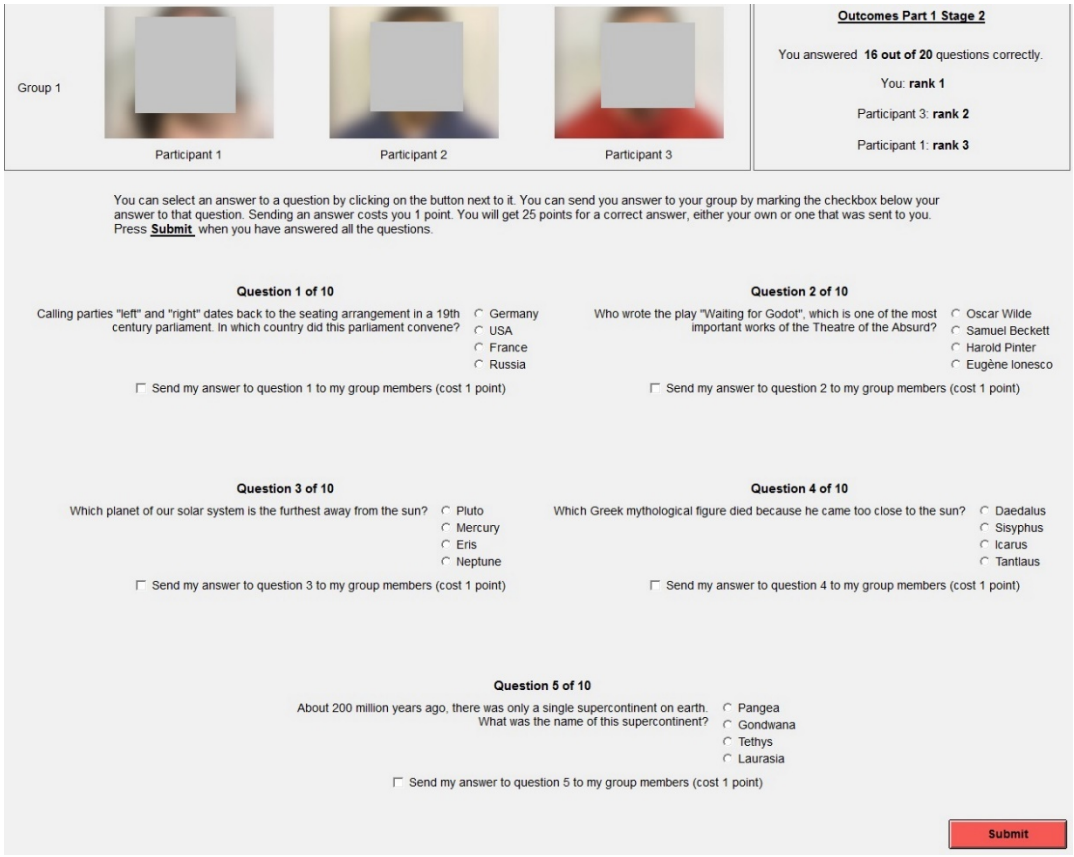
The Appendix contains the following additional tables:

1. Table A1 summarizes inference with Wilcoxon-Mann-Whitney tests on the treatment effect of rank feedback and relative pay on the willingness to help conditional on rank in the competition
2. Table A2 lists and presents summary statistics of all questionnaire items from which the indices on attitudes toward cooperation, competition, group work and autonomy are constructed

Figure A1. Information Provided at the End of Part 1

Notes. This figure displays the information that participants saw at the end of the timed general knowledge test in each experimental condition. The three screenshots to the left show the feedback screens in the baseline condition (top), the private rank feedback condition (middle) and the relative pay condition (bottom). The screenshot to the right shows the feedback screen in the public rank feedback condition. The faces of participants are not shown here to preserve their anonymity. They were shown when used in the experiment.

Figure A2. 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 questions 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 s/he had obtained an additional bonus payment.

Figure A3. Pictogram of the Inclusion of the Self in Other (IOS) scale

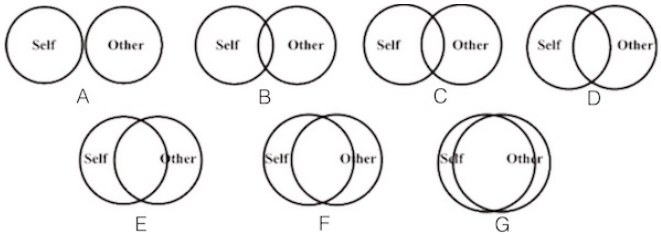


Figure A4. Empirical CDF of Performance on Timed General Knowledge Test by Condition

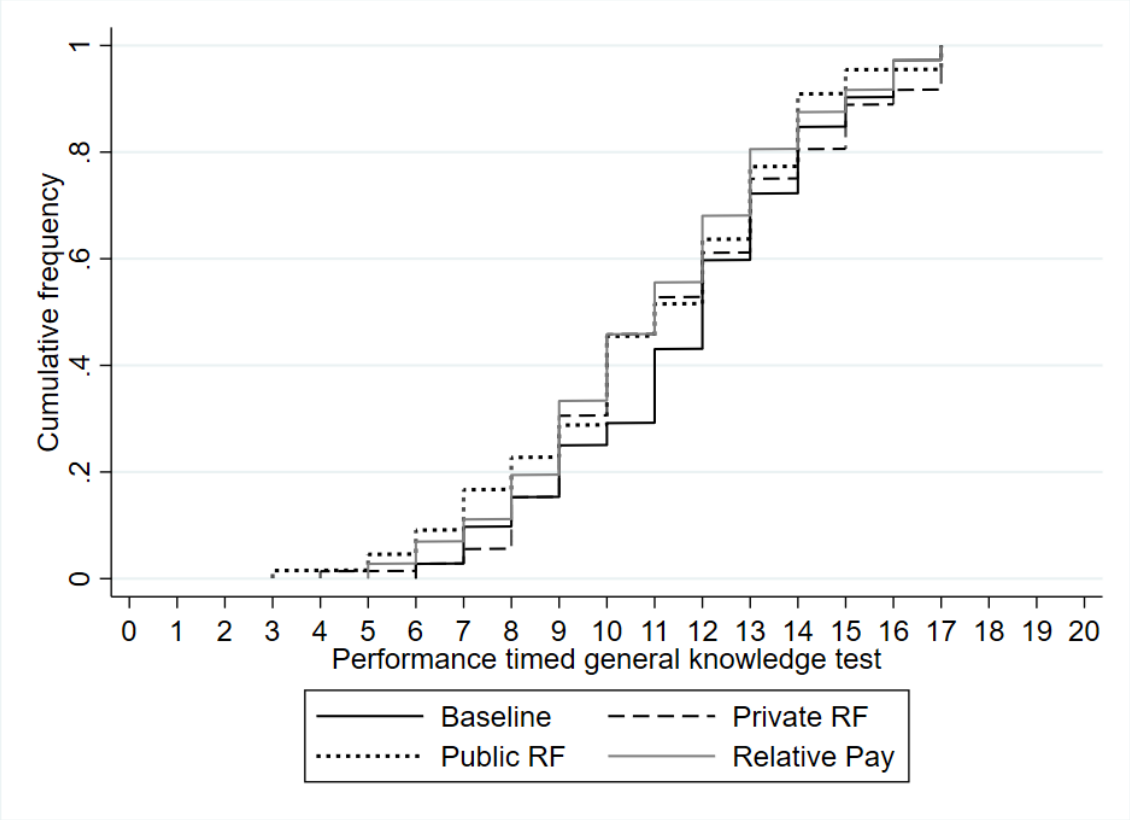


Figure A5. Help by Treatment and Rank in Competition (Means)

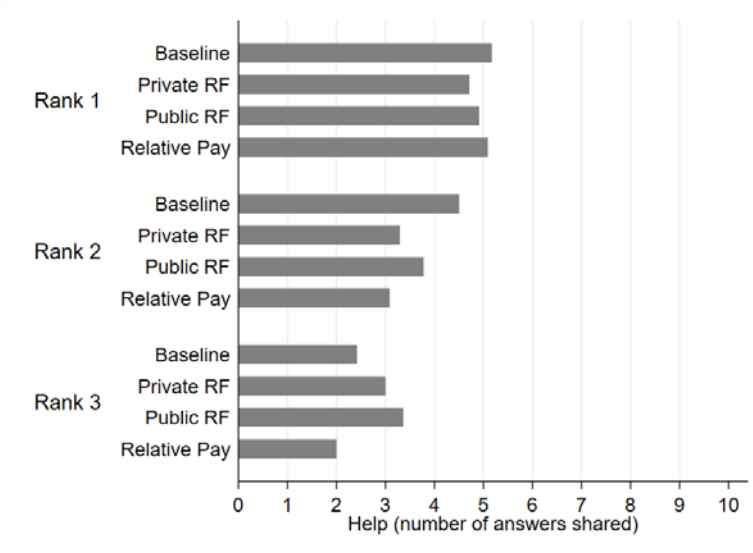


Table A1. Wilcoxon-Mann-Whitney Tests

rank	compare number of answers shared across conditions	p-value
1	Baseline-Private RF	0.6539
	Baseline-Public RF	0.8046
	Baseline-Relative Pay	0.9629
2	Baseline-Private RF	0.2167
	Baseline-Public RF	0.3028
	Baseline-Relative Pay	0.0836
3	Baseline-Private RF	0.6136
	Baseline-Public RF	0.7348
	Baseline-Relative Pay	0.4416

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 lists all the four 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.