Labour Economics 66 (2020) 101865

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Relative performance feedback to teams

William Gilje Gjedrem, Ola Kvaløy*

University of Stavanger, UiS Business School, 4036 Stavanger, Norway

ARTICLE INFO

Keywords:
Teams
Performance feedback
Performance pay
Experiment

ABSTRACT

Work teams often receive feedback on how well their team is performing relative to their benchmarks. In this paper, we investigate experimentally how teams respond to relative performance feedback (RPF). We find that when subjects work under team incentives, then RPF on team performance increases the teams' average performance by almost 10%. The treatment effect is driven by higher top performance, as this is almost 20% higher when the teams receive RPF compared to when the teams only receive absolute performance feedback (APF). The experiment suggests that top performers are particularly motivated by the combination of team incentives and team RPF. We also find notable gender differences. Females respond negatively to individual RPF, but even more positively than males to team RPF.

1. Introduction

People prefer high rank to low rank. Even when rank is independent from monetary outcomes, people are willing to take costly actions in order to climb the ladder. "….rank among our equals, is, perhaps, the strongest of all our desires" wrote Adam Smith in 1759. Modern organizations utilize this basic human insight by providing employees with feedback on their relative performance in order to motivate them to work harder.

However, although rank and relative performance feedback (RPF) is such a basic ingredient in competitive environments, it is only recently that scholars within economics have systematically studied how people respond to RPF. The early literature on relative performance evaluation studied the effect of connecting rank to monetary incentives (see Lazear and Rosen (1981) seminal contribution on rank order tournaments). More recent theories on competitive preferences and status concerns (Frank, 1985; Clark and Oswald, 1996; Auriol and Renault, 2008) suggest, however, that rank *per se* motivates effort.[1] It has now been demonstrated, through controlled experiments in the lab and in the field, that RPF indeed affects individual behavior, even when relative performance does not affect pay. For example, Blanes i Vidal and Nossol (2011), Kuhnen and Tymula (2012), Tafkov (2013), Charness et al. (2014), and Kramer et al. (2016) find performance improvements in situations where RPF is provided, while Hannan et al. (2008), Hannan et al. (2013), Azmat and Iriberri (2016), and Gjedrem (2018) find significant context specific effects of RPF. There are also studies that do not find any positive effects of RPF. Guryan et al. (2009), Eriksson et al. (2009) and Bellemare et al. (2010) find no significant effects, while Barankay (2012) find that removing RPF positively affected productivity.

Relative performance feedback is also provided to groups of individuals, like firms, or teams within firms, who compete against each other and receive feedback about their relative performance. Sales teams or R&D teams, for instance, are benchmarked against similar teams in other firms. Moreover, firms often set up internal competitions between teams in order to sell more or innovate more (Birkinshaw, 2001; Marino and Zábojnik, 2004; Baer et al., 2010). Successful teams are typically compensated by some monetary rewards, but team competitions *per se* may also be motivating. There are only a few studies on the effects of team RPF with real-effort task, see e.g. Delfgaauw et al. (2013).[2] However there is a larger literature on behavior in group contests using other types of tasks, see a recent survey by Sheremeta (2018) for details on this.

We contribute to the existing literature by investigating how teams respond to relative performance feedback while working on a real-effort task. We first build a model, considering how people's utility is affected

* Corresponding author.
E-mail addresses: william.g.gjedrem@uis.no (W.G. Gjedrem), ola.kvaloy@uis.no (O. Kvaløy).

[1] While status concerns may be independent from competitive preferences, the latter is often seen as a consequence of the former. People like to outperform others because it gives social status (see e.g., Charness and Grosskopf, 2001). We will use the two terms synonymously in this paper, and will not try to disentangle the two.

[2] They study competition between stores in a Dutch Retail chain and find that RPF to stores (i.e. teams) improves sales even when rank do not affect monetary outcomes.

https://doi.org/10.1016/j.labeco.2020.101865
Received 3 May 2019; Received in revised form 3 June 2020; Accepted 4 June 2020
Available online 1 July 2020

by feedback on rank and peer pressure. In particular, people may experience utility from being ranked against others, and from peer pressure of observing others' performance. From a theoretical perspective, one may think of at least two reasons why people might respond differently to team feedback compared to individual feedback. The first relates to status concerns and competitive preferences: The utility from winning together with a team might potentially be different from the utility of winning alone, since one then share the pleasure with other team members. Similarly, the costs of losing as a team might be different from the costs of losing alone, as one can find common support and comfort. The second relates to peer pressure and "team spirit". As demonstrated theoretically (Kandel and Lazear, 1992) and empirically (e.g., Babcock et al., 2015; Corgnet et al., 2015), peer pressure can motivate workers to exert effort in teams. Team-based incentive schemes may create peer pressure since low (high) effort has a negative (positive) externality on peers' pay. If peers also care about team rank, then team RPF may create additional peer pressure within the team.

We investigate RPF to teams by conducting a controlled laboratory experiment consisting of eight treatments. In each treatment, subjects work on a real-effort task for six periods. We primarily vary treatments along two dimensions: team or individual incentives, and team or individual feedback. However, to establish a "baseline" of performance, we also have treatments in which subjects only receive absolute performance feedback. Under RPF, individuals (teams) are always compared with two other individuals (teams), i.e. after each period, each individual or team is ranked as either number 1, 2 or 3. Each team consists of three subjects, so each subject earns one third of total team output when provided with team incentives. The monetary outcomes are independent from feedback rankings.

While our main objective is to investigate the effect of team RPF, our secondary objective is to understand when and why team RPF has an effect. It is difficult to disentangle between the two main mechanisms that could make people respond differently to team RPF compared to individual RPF (status concerns and peer pressure). Our approach is to remove (or at least reduce) peer pressure by letting people work *on behalf of* teams, where the others in the team do not work. We thus also ran two "team leader" treatments, where workers acted as team leaders and worked on behalf of their team.

In sum, these eight treatments can provide answers to our two main questions: How and why do individuals' performances in teams respond to relative performance feedback? The main results is that RPF to teams indeed increase performance, and that competitive preferences, rather than peer pressure or "team spirit", seem to drive the results. More specifically, our results can be summarized as follows: We find that when subjects are exposed to team incentives, then RPF on how their team is doing compared to two other teams increases the team's average performance by almost 10%. The treatment effect is driven by higher top performances.[3] The best performance within each team is on average almost 20% higher when teams receive team RPF comparted to when teams only receive team APF. These effects more or less disappear under individual incentives and individual RPF. Our experiment thus suggests that some subjects are particularly motivated by the combination of team incentives and team RPF. The strong effect on top performers, and the insignificant effect on other team members, indicates that team spirit is not a main explanation of our results. Our results from the team leader treatments support this conjecture. We find that team leaders receiving RPF perform significantly better than team leaders who only receive absolute performance feedback, indicating that status concerns or competitive preferences better explain our results than peer pressure or team spirit.

The positive effect of team RPF complements Delfgaauw et al. (2013) who in a field experiment find positive effects of team RPF under weak team incentives. In contrast to us, they do not compare with individual RPF, nor do they study interaction effects between team RPF and team incentives. Our results also complement van Dijk, Sonnemans, and van Winden (2001) findings that team incentives lead to higher top performances. In our experiment, team RPF is needed in addition to team incentives in order to improve top performance and thereby compensate for the drop in performance observed when simply moving to team incentives.

However, our results contrast with a field study by Bandiera et al. (2013). They find that ranking teams reduces overall performance, as lower ranked teams decreased their productivity. Our experiment has important differences though, as Bandiera et al. (2013) study this in a within-subject experimental design with endogenously formed teams, whereas we use a between-subject design with exogenously formed teams. They argue that the reduction in productivity is due to subjects starting to sort into teams based on ability rather than social ties, which leads to more free-riding in low productive teams with weaker social ties.[4] Hence, the reduction in performance in their study is likely to be driven by a factor not present in our study. Moreover, in a recent field experiment by Chen and Gong (2018), it has been shown that teams formed endogenously exerted higher effort than exogenously formed teams. This supports the idea that social ties within an endogenously formed team affect effort per se.

We also study gender effects. Previous literature has shown that gender is an important variable in order to understand competitive preferences (for an overview see Croson and Gneezy, 2009; Bertrand, 2011). In particular, females tend to shy away from competitive settings and they are more risk averse than males (see e.g., Niederle and Vesterlund, 2007; Charness and Gneezy, 2012). When faced with a competitive environment, males tend to respond positively, while females do not (Gneezy et al., 2003; Gneezy and Rustichini, 2004). Azmat and Iriberri (2016) also find that females are less responsive to individual RPF than males. Gender differences in response to team RPF have not been studied, but it has been found that women are less averse to competition if they compete as teams rather than as individuals (Healy and Pate, 2011; Dargnies, 2012; Flory et al., 2015). Moreover, a recent experiment by Kuhn and Villeval (2015) show that women are more likely than men to enter team-based environments. Our results complement this literature. . Indeed, we find that females respond negatively to individual RPF also in our study, but even more positively than males to team RPF. For males, team incentives have a strong negative effect compared to individual incentives, unless accompanied by team RPF. For females, incentives do not matter to the same degree. Team RPF has a strong positive effect regardless of the incentive system.

On a more general level, our results can contribute to explaining why team incentives are so common, despite the well-known free-rider problem. A majority of firms in the US and UK report some use of teamwork in which groups of employees share the same goals or objectives, and the incidence of team work and team incentives has been increasing over time (see e.g., Lazear and Shaw, 2007; Bandiera et al., 2013, and the references therein). Team incentives are puzzling because the individual incentive effect is quite small, and the temptation to free-ride on peers' effort is high (Holmstrom, 1982). Empirical research shows, however, that team incentives do surprisingly well, and it has been hard to actually identify strong free-rider effects.[5]

[3] We use the performance of subjects in the final work stage to categorize "top performance" and "top performers", i.e. the best subject within each team in the final stage is categorized as the "top performer" in that team, regardless of how they performed in the other working stages. Hence, the effect is driven by higher performance of the most productive individual in the final stage within each team, when comparing treatments with and without team RPF. We use top performers or best performers interchangeably throughout this article.

[4] It is argued by Bandiera et al. (2013) that social ties within the team ameliorate free-riding.

[5] A range of studies employing different empirical approaches have identified mixed effects of team incentives. In some field studies, there is an

Peer pressure and team spirit is a common explanation for why team incentives work better than standard theory predicts.[6] As Alchian and Demsetz (1972) notes "If one could enhance a common interest in non-shirking in the guise of team loyalty or team spirit, the team would be more efficient. The difficulty, of course, is to create economically that team spirit and loyalty". Theorists have also investigated more formally how firms can create the kind of team spirit that Alchain and Demsetz call for. Kandel and Lazear (1992) introduce a peer pressure function and discusses how firms can manipulate peer pressure by e.g. investing in team spirit building activities. Akerlof and Kranton (2000, 2005) incorporate identity into an otherwise standard utility function. They discuss how teams or firms can transform the workers' identity from "outsiders" to "insiders" by creating common goals that each individual shares with their team or firm.

Relative performance feedback to teams can be seen as a means of creating the kind of team spirit or identity discussed by these theorists. However, our results points to a different mechanism. Top performers respond strongly to relative performance feedback in our experiment, while the effect is insignificant for the other team members. Moreover, team leaders respond even when their peers do nothing. The theoretical framework we present indicate that our results are mainly driven by status concerns and/or competitive preferences rather than team spirit and peer pressure.

To the best of our knowledge, we are the first to study the effect of relative performance feedback to teams in a laboratory experiment using a real effort task. However, our paper relates to a larger literature studying how intergroup competitions or comparisons affect intra group behavior. Social psychologists have argued that intergroup comparisons can motivate group members to increase the contribution to their own group (Turner, 1975). A number of experiments have supported this conjecture. Group competition can induce more cooperation (Bornstein and Ben-Yossef, 1994), less free-riding (Bornstein et al., 1990; Erev et al., 1993; Bornstein and Ben-Yossef, 1994), and better coordination (Bornstein et al., 2002). See also a recent survey by Sheremeta (2018) on behavior in group contests and the references therein. Notably, Erev et al. (1993) find, using a field experiment, that prize competition between teams eliminates the free-rider effects of team incentives. We find a similar result, but with the important difference that our subjects compete without monetary prizes.

Some recent papers find that intergroup comparisons can improve intragroup contributions even without monetary prizes. Tan and Bolle (2007), Burton-Chellew and West (2012), and Böhm and Rockenbach (2013) find that subjects contribute more to a public good if their group's contribution is compared to another group.[7] This clearly resembles and supports our findings on team RPF, but there are significant differences. Importantly, we conduct a real effort experiment where subjects have to work on a specific task, in contrast to the public goods experiments (PGEs) where "effort" is a simple decision variable. Moreover, the experiments citied above do not study the interaction effects between different incentive regimes and different feedback systems, which is our focus.

The rest of the paper is organized as follows. In Section 2 we present our experimental design. In Section 3 we present a theoretical framework and provide some behavioral predictions. In Section 4 we present the results from the experiment, while Section 5 concludes.

2. Experimental design

2.1. Task

Subjects work on a real-effort task of decoding numbers into letters, used in several other related experiments (e.g., Charness et al., 2014). Specifically, subjects have a list of letters each assigned with a corresponding number, and the task is to decode given sequences of four numbers into their respective letter. The experimental session consists of six working stages, each lasting five minutes. There is a break in between each stage, and during the break subjects receive feedback (explained below). Participants earn a 100 NOK show-up fee ($1 ≈ 8 NOK). In addition, they can earn money by solving tasks, explained in the next subsection.

There are two main reasons why we have chosen this particular task. First, it requires no prior knowledge and is easy to understand. Second, we expect the task to be boring and tiresome, generating disutility of effort. To ensure disutility of effort we allow subjects to engage in alternative activities during the experiment, such as using their mobile phones for internet surfing. We require them to remain in their seat and refrain from communicating with other participants but tell them they can freely allocate their time to whatever suits them the most. Distracting activities are typically also present in the workplace so, if anything, these activities only make it more similar to the field. The task also provides a precise measure of output, which is our productivity indicator. Each session has the same sequence of number-decoding tasks. Subjects cannot proceed to a new task before the current task is correctly solved.

2.2. Treatments

We primarily vary treatments along two dimensions: team or individual incentives, and team or individual feedback. However, to establish a "baseline" of performance, we have two treatments in which subjects only receive absolute performance feedback. Feedback always concerns performance in the previous stage only, i.e. no aggregate information of multiple stages is displayed. In all treatments, subjects learn about their individual absolute performance. Moreover, in all team treatments, subjects learn the total absolute performance of their team. When subjects receive RPF, individuals (teams) are always ranked relative to two other individuals (teams), and they are ranked relative to the same individuals (teams) throughout the experiment (randomly assigned). Team members work independently on the tasks, and there are no complementarities in production. Teams also remain unchanged throughout the experiment (randomly assigned).

The piece-rate for a correctly solved task is 1 NOK. In the individual incentive treatments, subjects earn the piece-rate multiplied with total number of tasks they solve. In the team incentive treatments, subjects earn the piece-rate multiplied with one third of the total number of tasks the team solved, i.e. all team members earn the same. Hence, monetary outcomes only depend on the number of tasks subjects or teams solve, not on feedback ranks.

Treatment names are structured as follows: It first denotes whether feedback is absolute (APF) or relative (RPF), then whether there are

overall performance improvement of team incentives, relative to individual incentives or relative to an absence of team incentives, see e.g. Knez and Simester (2001), Hamilton et al. (2003) and Boning et al. (2007). On the other hand van Dijk et al. (2001), Vandegrift and Yavas (2011), and Chen and Lim (2013), using controlled laboratory experiments to study team incentives, do not find any overall change in performance. van Dijk et al. (2001) do find that some subjects improve, but this is offset by others who free-ride. Still others find a negative effect of introducing team incentives. Nalbantian and Schotter (1997) find extensive shirking behavior under different types of team incentives, but competition between teams for a fixed price increases performance significantly.

[6] It should be noted that there are not only so-called behavioral or non-monetary reasons why team incentives might work. Team incentives can exploit complementarities and foster cooperation (Holmström and Milgrom, 1990; Itoh, 1991, 1992; Macho-Stadler and Pérez-Castrillo, 1993; Büyükboyacı and Robbett, 2017; Büyükboyaci and Robbett, 2019). Team incentives can also be desirable in repeated settings, as it strengthens implicit incentives, see Che and Seung-Weon (2001) and Kvaløy and Olsen (2006). However, experimental investigation of team incentives, like the one present in the paper, abstract from such technological team effects.

[7] See also Sausgruber (2009) who does not find significant effects from intergroup comparisons.

Table 1
Summary of treatments.

Treatment	Feedback type	Compensation scheme	Level of feedback	Number of subjects	Number of clusters (independent observations)
RPF-ind-ind	Relative	Individual	Individual	51 subjects (20 females, 31 males)	17
RPF-ind-team	Relative	Individual	Team	45 subjects (14 females, 31 males)	5
RPF-team-ind	Relative	Team	Individual	63 subjects (27 females, 36 males)	21
RPF-team-team	Relative	Team	Team	54 subjects (27 females, 27 males)	6
APF-ind-ind	Absolute	Individual	Individual	68 subjects (29 females, 39 males)	68
APF-team-team	Absolute	Team	Team	57 subjects (23 females, 34 males)	57
APF-teamleader	Absolute	Team	Individual	93 subjects (50 females, 43 males)	31
RPF-teamleader	Relative	Team	Team	84 subjects (49 females, 35 males)	28

individual (ind) or team (team) incentives, and finally whether the level of feedback is individual (ind) or team (team) based.

We introduce treatments gradually. All treatments are summarized in Table 1 and then explained below.

We start by keeping one dimension fixed and explain treatments that contain RPF first.

In the *RPF-ind-ind* treatment, subjects earn individual incentives and receive individual RPF. The individual RPF consists of performance information about two other participants in the session. They learn how many tasks they have solved themselves, and their performance is ranked (from 1 to 3) and they are also informed about how many tasks the other two subjects solved. In addition to the show-up fee, subjects earn the piece-rate multiplied with the number of tasks they solve.

In *RPF-ind-team* treatment, subjects still earn individual incentives, but RPF is changed and now concerns teams rather than individuals. They learn how many tasks they have solved themselves, and the team RPF consists of performance information about their own team and two other teams in the session.[8] In particular, the team's performance is ranked (from 1 to 3) and they learn how many tasks the other two teams solved. In addition to the show-up fee, subjects earn the piece-rate multiplied with the total number of tasks they solve.

In the *RPF-team-ind* treatment, subjects still receive individual RPF, but incentives are changed and now concern team outputs rather than individual outputs. The individual RPF consists of individual performance information about themselves and the two other team members.[9] In particular, their performance is ranked (from 1 to 3) and they learn how many tasks themselves and the other two subjects solved. In addition to the show-up fee, subjects earn the piece-rate multiplied with one third of the total number of tasks their team solves.

In the *RPF-team-team* treatment, subjects receive both team RPF and team incentives, rather than individual RPF and individual incentives. They learn how many tasks they solved themselves, and the team RPF consist of performance information about their own team and two other teams. In particular, the team's performance is ranked (from 1 to 3) and they learn how many tasks their own team and the other two teams solved. In addition to the show-up fee, subjects earn the piece-rate multiplied with one third of the total number of tasks their team solves.[10]

Next, we introduce our "baseline" conditions, where we do the same variations as with RPF, only with APF instead of RPF.

In the *APF-ind-ind* treatment, subjects earn individual incentives and receive individual APF. Importantly, they do not learn anything about the performance of any others. In addition to the show-up fee, subjects earn the piece-rate multiplied with the total number of tasks they solve.

In the *APF-team-team* treatment, subjects earn team incentives and receive team APF, rather than individual incentives and individual APF. They learn how many tasks they solved themselves and the total of their own team.[11] However, they do not learn anything about the performance of any other teams. In addition to the show-up fee, subjects earn the piece-rate multiplied with one third of the total number of tasks their team solves.

The primary use of APF treatments is to establish some baseline performances. Thus, we have only included APF treatments that are of main interest to compare with RPF treatments. Notice also that all treatments actually include APF, and hence RPF is an additional piece of information in the RPF treatments.

Our theoretical framework that follows in the next section, propose two explanations as to why people respond more positively to team RPF: status concerns / competitive preferences and peer pressure. In an effort to disentangle these effects, we separately ran two additional "team leader" treatments, where subjects acted as team leaders and worked *on behalf of* their team.[12] In these "team leader" treatments we have reduced peer pressure, at least in terms of team spirit, since the others in the team do not work. We use the same setup as in the other treatments, and the only changes are explained below.

In the APF-*teamleader* treatment, subjects work on the task as the team leader. In the instructions, subjects are told that they have been selected as the team leader in a team of three subjects. During the breaks, they receive feedback only about the performance of the team leader.[13] Incentives are team-based: In addition to the show-up fee, all three sub-

[8] Subjects are only informed about the team performance of the other two teams, not about the team member's performances.

[9] We choose to provide intra group individual RPF to keep the setup somewhat realistic, see Kramer et al. (2016) for a similar treatment. An alternative would be to base the individual RPF on the performance of two randomly chosen subjects. However, in a team setting, this alternative is seldom seen in real workplaces.

[10] RPF-ind-ind and RPF-team-ind are referred to as individual RPF treatments, whereas RPF-team-team and RPF-ind-team are referred to as RPF treatments.

[11] As subjects learn about the performance of themselves and the performance of the team in total, there is some inherent information about relative performance as well. Knowing the total output of the team will enable to subject to consider whether the performance is above or below the average of the team, and an imprecise rank.

[12] These treatments were organized the following way: Subjects were told in the instructions that they had been picked to lead a team of three, and that the performance of the team depended only on their effort as the team leader. Unknown to the participants, all subjects were in fact assigned as a team leader. We then randomly matched all team leaders with two other passive members, to form all teams. Thus, each subject played two roles, both as a team leader and as a passive member of two other teams. Subjects only knew about their role as the team leader during the experiment, and were told afterwards that they had also been a passive member of two other teams. Hence, there was only one team leader per team. We paid subjects both for their effort as team leader, and for their role as team member in two other teams. Therefore, all information in the experimental instructions were true, and there was no deception (i.e. they were all real team leaders, and the team's performance and team payment depended solely on the performance of them as the team leader). We only omitted information about their role as passive members of two other teams until the end of the experiment.

[13] As all subjects worked as team leaders, this feedback was in fact only information about their own performance (as team leader). In other words, passive members of teams did not get continuous feedback on the performance of their team leader, but this was unknown to the participants. In the experimental instructions, we explicitly informed them that they would receive feedback about their performance as team leader, which they also did get.

jects in the team earn the piece-rate multiplied with one third of the total number of tasks their team leader solves.

In the RPF-*teamleader* treatment, subjects work on the task as the team leader. In addition to the feedback provided in the APF-teamleader treatment, they also receive team RPF. Specifically, during the breaks, the team leader's performance is ranked (from 1 to 3) against the performance of two other team leaders from two other teams, and they learn how many tasks they solved. Monetary incentives are the same as in APF-teamleader: In addition to the show-up fee, all three subjects in the team earn the piece-rate multiplied with one third of the total number of tasks their team leader solves.

In particular, the team's performance is ranked (from 1 to 3) and they learn how many tasks their own team and the other two teams solved.

In order to highlight the team leader role, and to minimize team spirit effects, we let the passive team members only see their team leader's performance at the end of the experiment, not during each break.[14] This also allowed for a simpler procedure: In each session, after all working periods, the team leaders were told that they have also been a passive member of two other teams. They then learned how much they had earned from their role as a team leader, and how much additionally they had earned as passive members of two other teams.

2.3. Procedures

The experiment was conducted at the University of Stavanger, Norway, in March and November 2015 and May 2017. We ran three sessions of each treatment over four days in March, except for the three sessions in RPF-ind-team that we ran in November.[15] The team leader treatments were conducted in May 2017. A session had up to 23 participants, and treatments with RPF or teams required a total number of participants that could be divided by three (and precisely 18 participants in RPF-ind-team and RPF-team-team). We recruited subjects through their student email accounts and posters on the University campus, and they signed up using the recruitment program Expmotor.[16] The student pool consists of a variety of students from three faculties: the faculty of Science and Technology, the faculty of Social Sciences, and the faculty of Arts and Education.[17] The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

We randomly seated subjects when they arrived in the computer lab. Each desk had a paper with written instructions, and we read the instructions aloud before the start of the experiment (instructions attached in the appendix). Then they worked on the task and received feedback during the breaks. Once the experiment concluded, we informed subjects about their total output and earnings. Then they completed a short questionnaire, where we asked for basic demographic details and elicited their ex post perceptions of the experiment. Specifically, we asked them how motivated they were to do the tasks, how they felt right now, and whether they thought the information in-between each stage affected them. They answered these questions on a scale from −5 to 5.

Each session lasted about 50 min. The average earnings for each participants were NOK 289 (about $35), which consisted of the 100 NOK show-up fee and the 189 NOK performance-related pay. A total of 515 subjects participated in the experiment. The uneven number of participants across treatments is due to overbooking and no shows.

3. Behavioral predictions

We will now present a simple theoretical framework enabling us to present some behavioral predictions, formulated as hypotheses. Our aim is not to test the presented model directly, but to provide a framework that enable us fix ideas and to disentangle the different mechanisms at play.

Let there be n agents in the economic environment. Agent i exerts effort e_i incurring a private cost $c(e_i)$ where $i = 1...n$, and where the cost function has standard properties $c'(e_i) > 0, c''(e_i) > 0$. He receives a wage $w(e_i, \dots, e_n)$ and is assumed to have the following utility function:

$$U_i = w(e_i, \dots, e_n) - c(e_i) + \theta v(e_i, \dots, e_n) - P(e_i, \dots, e_n)$$

The function v represents what we may call "rank utility", i.e. the utility from comparing performance with other agents. If agents have competitive preferences, they will enjoy outperforming others, but suffer from performing worse.[18] Building on Clark and Oswald (1996), we let the competitive preferences take the form $\theta v(e_i - e^*)$ where e^* is the benchmark to which the agents compare themselves (average performance in their model), and θ represents the weight the agent put on rank utility.[19] This weight can be interpreted as status concerns.

In addition, we add a peer pressure function P, similar to Kandel and Lazear (1992). Peer pressure is social and/or moral costs, for example disutility of being a free-rider, as functions of own and peers' effort. Like Kandel and Lazear, we assume that if an agent's effort has positive externalities in terms of increasing the other agents' utility, then $\frac{\partial P}{e_i} < 0$. In other words, agents can reduce peer pressure by increasing their own effort. However, peer pressure is also a function of peers' effort. For a given effort level from agent i, more effort from the peers increases peer pressure. This way, teams can generate "team spirit" by lifting each other's effort via peer pressure. Kandel and Lazear distinguish between shame and guilt, where shame is external pressure and guilt is internal pressure. With shame, the peer pressure costs are related to the other agents' observation of agent i's effort, while with guilt, the agents may feel peer pressure even if the other agents cannot observe their effort.

Let us first assume no peer pressure and no rank utility. Then individual incentives of the simplest type, $w = e_i$, clearly do better than team incentives $w = \frac{1}{n}\sum_{i=1}^{n} e_i$, since optimal effort is given by $1 = c'(e_i)$ and $\frac{1}{n} = c'(e)$, respectively. This is the classical $\frac{1}{n}$ free-rider problem. With respect to the treatments explained in the previous section, subjects in the APF-ind-ind treatment are expected to outperform subjects in the APF-team-team treatment.

Assuming no peer pressure or rank utility there should be no performance difference across treatments within the same incentive scheme.

[14] Admittedly, this was not made explicitly clear to the team leaders, so the team leader might have been under the impression that the passive team members got feedback about the performance of the team leader each period. However, this does not alter the basic rationale behind these two treatments, namely, to investigate subjects working *on behalf of* teams, and thereby disentangle team sprit from status concerns / competitive preferences. Further research could even try to disentangle the latter two by varying to what extent the passive team members can observe RPF.

[15] We have no reason to believe that the different month for this treatment would cause any differences per se, and predetermined characteristics of subjects participating in this treatment are very similar to the other treatments, as can be seen in the appendix Table A1.

[16] Developed by Erik Sørensen and Trond Halvorsen at the Norwegian School of Economics (NHH).

[17] About 47% were females, and the average age was just above 25 years. Slightly more than half of the participants were Norwegian citizens. About 13% of all participants were students from an Economics program.

[18] This general specification of rank utility opens for a set of underlying preferences. Rank utility may for instance include social preferences such as envy. Moreover, one could also include wage in v , rather than just effort. However, the general specification also opens for the possibility that rank utility is affected by wage via effort, as effort affects wage. Hence, it may be that, under performance related payment, rank utility follows from wage differences. Finally, absolute performance feedback may itself may provide utility, which we abstract from in our model.

[19] The main difference from Clark and Oswald (1996) is that there is more information with direct specific rank compared to just an average, which in our experiment could provide an even stronger impression on agents who receive this information.

The optimal solution for the individual is to maximize own payoff, which happens without any concerns about the performance information.

Relaxing these assumptions, we first allow peer pressure to affect the performance. By introducing peer pressure in teams, the free-rider problem can be reduced. Under team incentives, the lower effort from agent i, the lower wage to the other agents in the team. If this has a personal cost for agent i, then $\frac{\partial P}{e_i} < 0$. Optimal effort is then given by $\frac{1}{n} - \frac{\partial P}{e_i} = c'(e)$ and will thus increase effort compared to the case without peer pressure. Whether or not team incentives do worse than individual incentives now depends on the strength of the peer pressure compared to the size of 1/n free-rider problem.

Assume also that agents have competitive and social preferences. If agents only get information about their own performance, then we can assume $v = 0$ for all effort levels. However, with relative performance feedback (RPF), then $\frac{\partial v}{e_i} > 0$ and hence RPF motivates effort. If we use the form $v(e_i - e^*)$ then feedback on team level (team RPF) would yield $v(\sum e_i - e^*)$ where $i = 1...t$ and t is the number of agents in the team, while e^* is the average performance of other teams. Given this specification, then cet. par. the motivational effect from RPF ($\frac{\partial v}{e_i}$) is the same for team RPF and individual RPF.

Both peer pressure and rank utility are expected to increase the overall performance of subjects. For peer pressure, subjects should increase effort to avoid the social and/or moral costs of low performance. For rank utility, people get utility from ranking above others, and disutility for ranking below. In both cases the feedback information foster competition, in which subjects would want to either maintain their position or improve on their relative position. Thus, given that both peer pressure and rank utility are expected to improve the overall performance of subjects, we expect the performance of subjects in treatments with any form of RPF to exert higher effort relative to APF (within the same incentive scheme). In terms of treatments introduced in the previous section, subjects in RPF-ind-ind and RPF-ind-team are expected to perform better than subjects in APF-ind-ind, and subjects in RPF-team-team and RPF-team-ind are expected to perform better than subjects in APF-team-team.

Hypothesis 1. ***(H1):***

A) *Under team incentives, subjects who receive any form of RPF perform higher than subjects who only receive APF.*
B) *Under individual incentives, subjects who receive any form of RPF perform higher than subjects who only receive APF.*

An interesting question is how feedback and incentives interact. Can team RPF work better under team incentives and vice versa? There are two potential mechanisms creating positive interaction effects. The first is via peer pressure: When agents are exposed to both team incentives and team RPF, peers suffer a double utility loss of low effort from agent i: lower team pay and lower rank utility. If the agents have (standard) concave utility functions over rank and wage ($v'' < 0$ and/or $u''(w) < 0$), then the marginal positive effect of effort from agent i on the agent j's utility is higher when the agents have *both* team incentives and team feedback, compared to when only one of the features is in place. The second mechanism is via status concerns. Agents may potentially put different weight on v when it is about team comparisons rather than individual comparisons, as sharing success with others or coping with a loss by finding common support and comfort may differ from individual success or loss. If this difference is a function of incentives, i.e. if agents put higher weight θ on rank v under team RPF when agents also are exposed to team incentives, then we have positive interaction effect.

The four RPF treatments vary along two dimensions, individual or team incentives and individual or team RPF. Hence, with individual incentives and individual RPF (RPF-ind-ind) as the starting point, we can gradually introduce team incentives (RPF-team-ind) and team RPF (RPF-ind-team) and finally both team incentives and team RPF (RPF-team-team) to capture any interaction effect.

Hypothesis 2. ***(H2):*** *There is a positive interaction effect between team incentives and team RPF.*

In our framework, heterogeneous responses to RPF can also give insight into whether status *per se* plays a role. Given our specification, unobserved ability differences should put more peer pressure on low ability workers. Hence, team RPF should potentially have a stronger effect on low performing agents if peer pressure is important. Moreover, differences in ability and/or performance within a team does affect rank utility v in our specification. Hence, if one observes higher team RPF response from the top performers within teams, the plausible explanation would be that the weight on status concerns, θ, differs between agents. Given the experimental design, we can compare the performance of the least productive and most productive agents across treatments.[20] Within the same incentive scheme, we expect low (high) performing agents in the team RPF treatments to outperform low (high) performing agents in both APF and individual RPF treatments.

Hypothesis 3. ***(H3):*** *Within the same incentive scheme, low (high) performing subjects in team RPF outperform low (high) performing subjects in the other treatments.*

Previous research on gender differences has suggested that gender may be an important variable to explaining different responses to incentives and information, as noted in the introduction. In particular, females may have weaker competitive preferences, and thus put less weight on θ in our framework. Indeed, related research by Azmat and Iriberri (2016) do find that the response to individual RPF is entirely driven by males in their setting. By the introduction of team RPF, however, this may change. There are empirical evidence suggesting that females are less averse to competition in teams (Healy and Pate, 2011; Dargnies, 2012) and that they are more likely to enter into team-based environments than males (Kuhn and Villeval, 2015). Importantly, Flory et al. (2015) find that hiring into teams with competition-based pay substantially attenuated the gender differences they observed when hiring into individual competition-based pay. This may imply that they put more weight on θ in a team-based environment. In turn, this may lead to less or no gender difference in performance with teams. Hence, we only expect males to respond to individual RPF, in line with previous findings. Moreover, within the same incentive scheme, we expect females to respond more positively to team RPF compared to APF and individual RPF.

Hypothesis 4. ***(H4):*** *Females in treatments with team RPF outperform females in other treatments.*

Recall that there are two reasons why team RPF may have a different effect than individual RPF in our framework. First, as noted above, agents may potentially put different weight on v when it is about team comparisons rather than individual comparisons i.e. θ may be different under individual RPF, compared to team RPF. Second, peer pressure works with team RPF, also in the absence of team incentives. If peers care about rank utility, then there are positive externalities from effort even without team incentives, and hence $\frac{\partial P}{e_i} < 0$, under team RPF. In other words, team RPF *per se* can create peer pressure.

While the latter effect (peer pressure) makes team RPF stronger than individual RPF, the former (status) can go both ways. The extent that status *per se* plays a role in team settings can be investigated by studying teams with no peer pressure. It might not be fully possible to remove all peer pressure, but it is natural to assume that the lower the peers' effort, the lower is the peer pressure to work hard, at least in terms of peer pressure as defined by Kandel and Lazear (1992). Hence, if status matters, then team RPF may work well even if the other agents do not exert effort at all. If this is the case, team RPF may be efficient also

[20] As we do not have an ex ante measure of ability, their positioning in the performance distribution may be correlated with the type of feedback they receive.

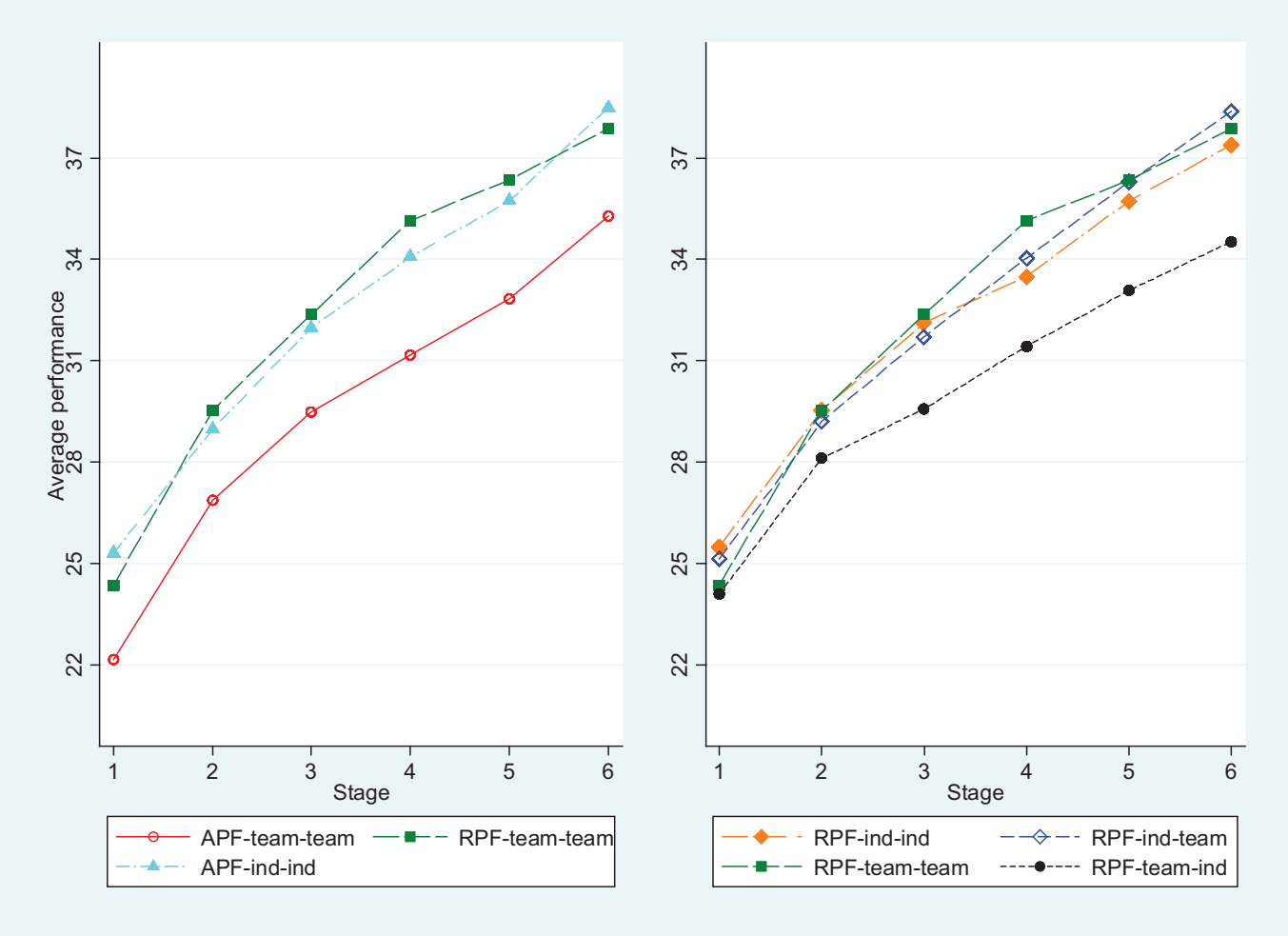

Fig. 1. Average performance across stages.

when the agent works on behalf of the team (as, say, team leader) and not only when he works along with other team members. This is what we seek to investigate in the team leader treatments; we aim to reduce peer pressure, in order to see if there are still differences in performance between team APF and team RPF.[21]

Hypothesis 5. ***(H5):*** *Team leaders who receive team RPF outperform team leaders who receive team APF.*

4. Experimental results

In this section, we present our experimental results. Fig. 1 intends to give an overview of the average performance of subjects in each treatment across all stages. As can be seen, the average performance increases throughout the experiment, likely to be driven by expected learning effects. The decoding scheme remains unchanged across all stages, and thus subjects are likely to memorize more decodes as stages go by. The performances are very similar across treatments, but the performances of subjects in APF-team-team and RPF-team-ind are notably lower than the performances of subjects in the other treatments. This will be more formally analyzed below.

4.1. Main treatment effects: non-parametric tests and regression analysis

We use the non-parametric tests Mann-Whitney *U* test (MW) and Randomization test (RT)[22] when comparing means throughout this section, besides when referring to regression analysis. In the non-parametric tests, we use the subject's average output whenever we refer to performance across all stages.[23]

The regression analysis provides a more formal test of differences across treatments, in which we can also control for other potentially important characteristics.[24] Reported in Table 2 are OLS and Random Effects GLS estimations, controlling for other factors such as age and

[21] Importantly, when other agents do not exert effort at all, the moral cost of underperforming relative to others in the team or free-riding is no longer relevant. Despite this, there might still other types of peer pressure present, for example, that earnings now depend entirely on the performance of the team leader.

[22] The Stata program *permtest2* by Kaiser (2007) is used to conduct the Randomization tests. This test is a powerful alternative to Mann-Whitney U-test, and is included to show that our estimates are robust to two different non- parametric estimation approaches. Several researchers have recently discussed the use of Randomization test in experimental papers (e.g., Imbens and Rubin, 2015; Young, 2018) as an important complement to the analysis. The RT-tests are based on 200.000 simulations.

[23] We do not use a cluster version of MW.

[24] In the appendix, Table A-, we check for randomization across treatments. Some minor differences exist, so controlling for such differences may prove important to the robustness of our findings.

Table 2
Main results: treatment effects on productivity.

Stage(s):	1st stage (1)	2nd stage (2)	All stages (3)
APF-team-team	Ref.	Ref.	Ref.
APF-ind-ind	3.147***	2.201***	2.582***
	(0.8371)	(0.3387)	(0.5722)
RPF-ind-ind	3.668**	2.717***	3.595***
	(1.3137)	(0.4953)	(0.8358)
RPF-ind-team	3.622***	2.863***	3.917***
	(1.2263)	(0.8106)	(0.6737)
RPF-team-team	2.639	2.734**	3.525***
	(1.8604)	(1.1023)	(1.3004)
RPF-team-ind	2.455*	1.527**	1.637**
	(1.3004)	(0.6038)	(0.6558)
Stage $_t$			2.384***
			(0.0772)
Constant	31.524***	35.125***	32.546***
	(3.0159)	(2.6259)	(2.6870)
Adjusted R^2	0.095	0.059	
Number of clusters	18	18	18
Observations	338	338	2028

Notes: OLS coefficients reported in columns (1) – (2) and Random Effects GLS coefficients reported in column (3), with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$.

gender.[25,26] APF-team-team is the baseline (reference group). We include a column for the 1st stage, the 2nd stage, and a column of all stages (the remaining stages are in the appendix, Table A2). The 1st stage is a "kick-off" stage, as any treatment effect of RPF is driven by the knowledge about future feedback, and not a response to the feedback itself (as found in e.g., Blanes i Vidal and Nossol, 2011). The 2nd stage is the first working stage after any feedback is provided, and the cleanest way to identify any treatment effects of RPF. We use multiple observations per subject whenever we refer to all stages in the regression analysis (i.e. one observation per subject per stage).

4.1.1. Performance differences across individual and team incentives

Consider first the performance differences across individual and team incentives, when subjects only receive APF. Comparing the two APF treatments,[27] the average performance in APF-ind-ind (32.4) is significantly higher (MW: $p = 0.01$, RT: $p = 0.01$) than in APF-team-team (29.6).[28] This is supported by the regression analysis in Table 2; see in particular the highly significant coefficient of APF-ind-ind in column (3). In other words, in the absence of relative performance feedback, performance is higher under individual incentives than team incentives. This could be a result of the free-rider problem discussed in Section 3, as subjects working under individual incentives solve, on average, almost 10% more tasks than those working under team incentives. This is also consistent with previous empirical findings of free-riding activity in teams (see e.g., Corgnet et al., 2015).

4.1.2. The effects of relative performance feedback on performance under team incentives

Under team incentives, any effects of RPF on performance will appear as a difference in performance across APF-team-team and the two treatments RPF-team-team and RPF-team-ind. We compare the performance across RPF-team-team and APF-team-team first. The average performance in RPF-team-team (32.6) is significantly greater (MW: $p = 0.09$, RT: $p = 0.02$) than in APF-team-team (29.6).[29] The regression analysis in Table 2 supports this finding, as the coefficient RPF-team-team is positive and highly significant, see in particular column (3). The effect is consistent throughout all working stages; see Table A2 in the appendix for details on all stages. This difference can also be observed in the left graph of Fig. 1. The performance is about 10% higher in RPF team-team compared to APF-team-team. The effect seems to be present from the very beginning of the experiment, suggesting that knowledge about the future performance feedback *per se* is enough to induce subjects to exert higher effort.

Next, we consider the performance of subjects across APF-team-team and RPF-team-ind. The non-parametric tests show no significant difference in overall performance across these treatments, see Table A4 in the appendix. However, the regression analysis in Table 2 suggests that subjects exert higher effort in RPF-team-ind relative to the baseline. This difference weakens in the final stages of the experiment, as can be seen in Table A2. Thus, there are suggestive evidence of a treatment effect also for individual RPF, but clearly weaker than for team RPF.

Result 1: *Under team incentives, performance is higher with both team RPF and individual RPF relative to team APF. This supports H1 (A).*

Finally under team incentives, we compare the performance in RPF-team-team to RPF-team-ind. Non-parametric tests show that the performance in RPF-team-team is significantly higher than in RPF-team-ind, but only from stage 2 and onward (MW: $p = 0.09$, RT: $p = 0.03$).[30] Regression analysis find no overall difference when stage 1 is included (Table 2), but when the analysis is done for stages 2–6 only the difference is significant ($p = 0.046$, see also Table A5 in the appendix).This difference can also be observed in the right graph of Fig. 1, from stage 2 and onwards. As the difference only exists from stage 2, it suggests that this is due to differences in the response to the content of the feedback provided in stage 1.

[25] In the regressions, we use robust standard errors clustered on sessions. However, as the number of clusters is low, it could downward bias standard errors. Therefore, we use a more conservative approach of only having (C-1) degrees of freedom when stating p-values, where C is the number of clusters.

[26] Alternatively, we could increase the number of clusters by applying the second highest level of clusters. This is at the level where teams receive feedback relative to two other teams in the team RPF treatments, i.e. nine subjects "interact" and must be part of the same cluster. For the other treatments, the level of interaction is at either three subjects or only one subject. Thus, in order to get a common level of clusters, we constructed quasi clusters of nine subjects for these treatments as well. This means that not all subjects within a quasi-cluster interact with each other, but all that do interact are certainly part of the same cluster. This approach only provided marginal differences to the results presented in the paper. The only part with notable differences is Section 4.2, where significance levels drop to 5% level or 10% level. For this approach in the analysis of gender, the interaction between team RPF and team incentive no longer remain significant for males, and the other variables drop slightly in significance.

[27] Strictly speaking, changing from individual to team incentives and from individual to team APF is a multiple change of conditions. However, there is no realistic middle way of only changing incentives or only changing to team APF.

[28] See Table A3 in the appendix. An alternative approach is to use team average rather than subjects' average. This also provides an overall significant difference with $p = 0.090$. The difference in 2nd stage (29.0 vs. 26.9) is also significant, MW: $p = 0.02$ and RT: $p = 0.02$.

[29] See Table A4. Using team average, provides an even more overall significant difference with $p = 0.026$ (based on 38 observations). The difference (26.9 vs. 29.5) in the 2nd stage is also significant, MW: $p = 0.04$ and RT: $p = 0.01$.

[30] Including the first stage leads to an insignificant difference (MW: $p = 0.13$, RT: $p = 0.05$), but considering the development in performance seen in Fig. 1, it is more appropriate to compare performance from stage 2 and onwards, especially if we want to capture the reactions after they observe feedback.

4.1.3. The effects of relative performance feedback on performance under individual incentives

Under individual incentives, the non-parametric tests do not show any difference in performance across treatments. The average performance in RPF-ind-ind (32.3) is not statistically different (MW: $p = 0.42$, RT: $p = 0.91$) from the performance in APF-ind-ind (32.4), see Table A6 in the appendix. Moreover, the average performance in RPF-ind-team (32.5) is not statistically different (MW: $p = 0.83$, RT: $p = 0.98$) from the performance in APF-ind-ind (32.4). Regression analysis supports the effect of no effect of individual RPF. However, there are suggestive evidence of a treatment effect of team RPF. The coefficient of RPF-ind-team is significantly higher than the coefficient for APF-ind-ind ($p = 0.015$) across all stages, see column (3) of Table 2. This difference, however, appears only in the final three working stages of the experiment.

Result 2: *Under individual incentives, performance is higher with team RPF, relative to individual APF. This partly supports H1 (B).*

Hence, the overall positive effect of team RPF seems to apply under both incentives, although less prominently under individual incentives. This may suggest the presence of peer pressure in team RPF treatments, especially given that the difference only appear in the final part of the experiment under individual incentives. Moreover, individual RPF does not seem to improve performance under individual incentives relative to individual APF. This is somewhat surprising. Given both the theoretical predictions of a positive effect from RPF and previous empirical findings from related research, we would expect to find such a performance difference. However, as reviewed in Section 1, there are also studies who have not found such an effect of RPF (e.g. Eriksson et al., 2009).

An interesting comparison, although a change of multiple conditions, is to compare the average performance of subjects in APF-ind-ind (32.4) to RPF-team-team (32.6). Statistical tests reveal no significant performance difference between them (MW: $p = 0.65$, RT: $p = 0.89$), see also Table A7. Hence, moving from APF-ind-ind to APF-team-team (step 1) revealed a performance decrease, which may indicate a free-rider problem. Moving from APF-team-team to RPF-team-team (step 2) revealed a positive effect of team feedback. The net result of these two steps cancel each other out, so that the addition of the team RPF (step 2) seems to offset the performance decrease observed when moving from individual to team incentives (step 1).

From Table A5, columns (1)–(3), we see that the effects discussed above are persistent throughout the working stages, and notably that subjects in RPF-team-ind do not perform any better than the baseline if the first stage is excluded.

4.2. Interaction effects

Consider now H2 and the four RPF treatments in a 2 by 2 design, varying between individual incentives or team incentives and individual RPF or team RPF (see Table 1).[31] In order to study how team incentives and team RPF affect each other, we employ a regression with an interaction term between team incentives c and team RPF r. This gives the following model:

$$y_i = \alpha + \beta_1 c_i + \beta_2 r_i + \beta_3 c_i r_i + controls + \varepsilon_i,$$

where $c_i = 1$ if subject i is working under team incentives (i.e., RPF-team-team or RPF-team-ind), and 0 if subject i is paid individual incentives; $r_i = 1$ if subject i is provided with team RPF (i.e. RPF-ind-team or RPF-team-team), and 0 if subject i is provided with individual RPF. Controls are the same as indicated in Table 2. Then β_1 is the effect on performance (y_i) of team incentives without team RPF, β_2 is the effect of team RPF without team incentives, while β_3 estimates the interaction between them.

In Table 3, we can see that there is a strong negative effect of team incentives alone, whereas team RPF alone has no significant effect. The net effect of both team incentives and team RPF is slightly positive, although not significant. However, we find a strong and positive *interaction effect* between team incentives and team RPF. This suggests that team feedback and team incentives complement each other, i.e. providing team RPF positively strengthens the influence of team incentives, and vice versa. This may be a result of reinforced marginal effect of effort for each subject when both incentives and team RPF are in place, and/or that the effects of team comparison is a function of incentives leading subjects to put a different weight θ on rank v.

Result 3: *There is a positive interaction effect between team incentives and team RPF. This supports H2.*

4.3. Heterogeneous effects

We use several approaches to investigate heterogeneous responses to performance feedback.

We start by categorizing each subject within a team as either best, worst or neutral, based on their performance in the final stage only.[32] In Fig. 2, we display the difference in performance between the best and the worst within each team across all stages. It shows a substantially larger performance gap between the best and the worst performance within each team in the RPF-team-team compared to any other treatment.[33] In Fig. A1 in the appendix, we see that high performers in the RPF-team-team treatment drive this difference.

Next in Table 4, we include a dummy variable (BiT) that takes the value of 1 whenever the subject is categorized as "best" (0 otherwise). This variable is then interacted with each of the treatments. The sum of the coefficients BiT and [treatment x BiT] is the number of additional tasks she solves relative to the two others within the team. APF-team-team is the reference group. Thus to compare their performance across treatments, say between best performers in RPF-team-team and APF-team-team (the baseline), one has to take the difference between them. That is, for the concrete example, one has to sum the coefficients for RPF-team-team and RPF-team-team x BiT in order to find the corresponding estimated difference.[34,35]

Consistent with Fig. A1, the best performers in RPF-team-team do significantly better than the best performers in the baseline ($p < 0.01$) and RPF-team-ind ($p < 0.05$). Notice also that the top performances in both treatments with individual incentives are similar to the top performance of the baseline (i.e. the performance difference is driven by the two other subjects in the team).

In Table 5, we run quantile regressions to explore further the heterogeneous effects across treatments. In these regressions, we use the average performance of each subjects as dependent variable rather than the fixed categorization of each team member used above. The quantile

[31] Recall that the reference for comparison is not exactly the same for subjects in the two different individual RPF treatments, as subjects in RPF-ind-ind are compared to two other subjects in the session, whereas subjects in RPF-team-ind are compared to two other subjects within the same team. One way to address whether this difference affects results is to compare within-team heterogeneity in performance across treatments, i.e. to compare variance within teams in RPF team-ind with variance within quasi teams in RPF-ind-ind. It turns out that this variance do not differ significantly (using Levene's robust test statistic (W_0) for the equality of variances).

[32] This means that a subject who is categorized as "best" due to her performance in the final stage, keep her categorization throughout all stages even though she might have been "neutral" or worst" in a previous stage.

[33] Notice that we have also included the RPF-ind-ind for comparison and constructed these "teams" based on the same subjects as their comparison group of two other subjects.

[34] Similarly, to compare the best performer in RPF-team-team to RPF-team-ind, the difference between them is the sum of the coefficients (RPF-team-team + RPF-team-team x BiT) – (RPF-team-ind + RPF-team-ind x BiT.

[35] Notice that when we interact the BiT variable with the treatment dummies, the total number of observations in these cells become one third of all subjects in that treatment, consequently reducing the statistical power.

Table 3
Changing incentives and feedback.

Stage(s):	All stages (1)	Stages 1–3 (2)	Stages 4–6 (3)
Individual incentives and individual RPF	Ref.	Ref.	Ref.
Team incentives	−2.427***	−2.372***	−2.483***
	(0.6124)	(0.5529)	(0.7384)
Team RPF	−0.192	−0.911*	0.526
	(0.5348)	(0.4729)	(0.7219)
Team incentives x Team RPF	2.854***	3.235***	2.473*
	(1.0557)	(0.9617)	(1.2894)
Stage $_t$	2.326***	3.315***	1.728***
	(0.1032)	(0.1837)	(0.1469)
Constant	33.002***	28.884***	38.137***
	(3.0820)	(2.5062)	(3.9754)
Number of clusters	12	12	12
Observations	1278	639	639

Notes: Random Effects GLS coefficients reported, with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

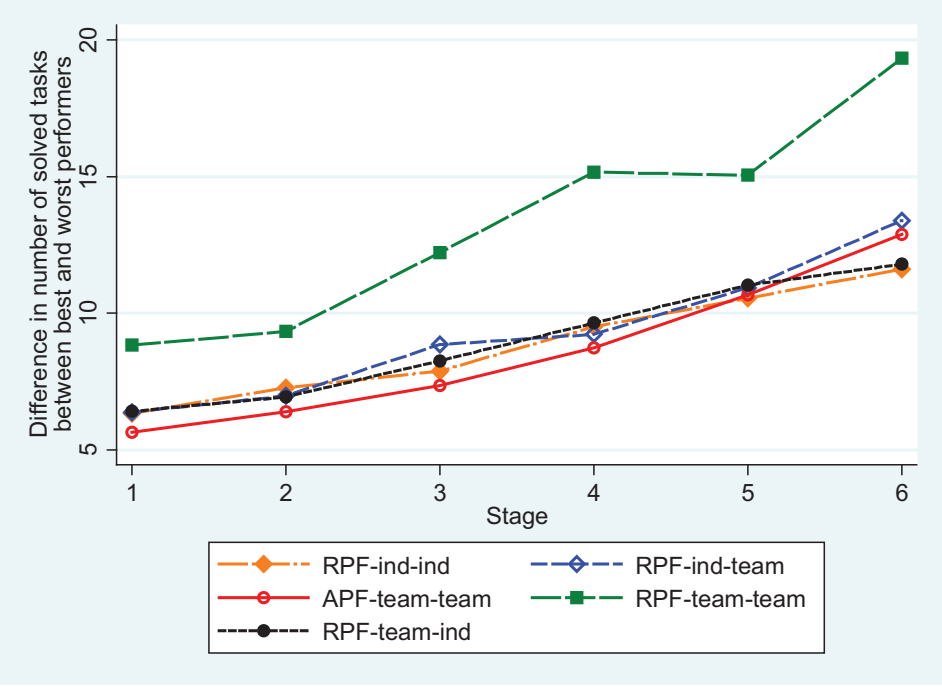

Fig. 2. Difference between best and worst performers across treatments.

regressions support previous findings in that the highest performances in RPF-team-team are substantially larger than the other treatments with team incentives. Moreover, we find a similar performance difference for the highest performers in the other team RPF treatment. Specifically, high performing subjects in RPF-ind-team outperform high performing subjects in APF-ind-ind ($p < 0.05$). As this difference is not observed in Fig. 2 or Table 4, it must be that there are multiple high performing subjects within the same team. Hence, in RPF-team-team, the performance seems to be driven by the high performing individual within each team, whereas in RPF-ind-team it seems to be driven by high performing teams.

On the other hand, there are no differences in performance amongst the lowest performing subjects across all treatments.

Result 4: *Within the same incentive scheme, high performing subjects in team RPF treatments outperform high performing subjects in APF treatments. There are no performance difference for low performing subjects. This partly supports H3.*

Result 4 suggests that subjects' weight on status concerns, represented by θ in the theoretical framework, differ between the agents, as we observe higher top performances in teams with team RPF compared to APF and individual RPF (at least under team incentives). Moreover, it suggests that peer pressure is not as influential in this setting, as this would imply a stronger response from low performing subjects. The result also illuminates previous findings showing that high performers are more willing to join teams (Hamilton et al., 2003) and less prone to free-ride under team incentives (van Dijk et al., 2001). In contrast, however, highly motivated subjects preferred to stand alone in an exper-

Table 4
Best performers across treatments.

Stages:	1st stage (1)	2nd stage (2)	All stages (3)
APF-team-team	Ref.	Ref.	Ref.
RPF-ind-ind	3.460*	2.358***	3.379***
	(1.6689)	(0.7359)	(0.8344)
RPF-ind-team	3.532**	2.675**	3.800***
	(1.5253)	(1.0310)	(0.9456)
RPF-team-team	1.697	1.598	1.574
	(1.9698)	(1.0318)	(1.3163)
RPF-team-ind	2.443	1.273	1.738**
	(1.5735)	(0.7484)	(0.7080)
BiT (Best in Team)	4.220***	4.393***	5.821***
	(0.9700)	(1.0025)	(0.8493)
RPF-ind-ind x BiT	0.979	1.181	1.348
	(1.8523)	(1.8428)	(1.7700)
RPF-ind-team x BiT	0.339	0.397	1.060
	(1.5031)	(1.5728)	(1.4278)
RPF-team-team x BiT	3.090*	3.534**	5.266***
	(1.4490)	(1.3368)	(1.4261)
RPF-team-ind x BiT	0.071	0.674	0.652
	(1.1530)	(1.1299)	(0.9382)
Stage $_t$			2.349***
			(0.0863)
Constant	26.744***	30.426***	26.907***
	(3.1055)	(2.4240)	(2.6755)
Adjusted R^2	0.247	0.273	
Number of clusters	15	15	15
Observations	270	270	1620

Notes: OLS coefficients reported in columns (1) – (2) and Random Effects GLS coefficients reported in column (3), with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks. BiT is a dummy variable taking value 1 if the subject is the best performer in his or her team, 0 otherwise. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

iment where effort was decided by choice, as weakly motivated subjects had an incentive to free-ride on highly motivated players (Herbst et al., 2015).

An interesting observation is that RPF seems to induce some subjects to perform better than they would have done if only provided with APF. This is the case in all RPF treatments, except when subjects were compensated as a team and received individual RPF. Hence, team incentives seems to crowd out the additional motivation that potential high performing subjects seem to get from RPF, but only when this feedback is individually based. Once RPF is team based, high performing subjects increases their performance substantially compared to the highest performing subjects in the baseline.

4.4. Gender analysis

The forth hypothesis concerns gender differences in response to performance feedback. In Table 6 we add a gender indicator and interact this with each treatment. We start the analysis by looking at differences across treatments for the same gender. Males in APF-team-team are the reference group. Under individual incentives, males in RPF-ind-ind outperform males in APF-ind-ind ($p < 0.05$), suggesting a motivational effect of individual feedback. Under team incentives, males in RPF-team-team ($p < 0.01$) and RPF-team-ind ($p < 0.01$) outperform males in APF-team-team. There are no differences between females across individual incentives treatments. Under team incentives, females in RPF-team-team and RPF-ind-team (both $p < 0.10$) outperform females in APF-team-team. See also Fig. A2 in the appendix. Consider next gender differences within the same treatment. The only difference we find is that males strongly outperform females in RPF-ind-ind ($p < 0.001$). The result on the effects of individual RPF on performance is consistent with previous empirical findings, in that males respond more than females to competition in general (Gneezy et al., 2003; Gneezy and Rustichini, 2004) and RPF in particular (Azmat and Iriberri, 2016).

Next, we use the first stage as control, and study how the performance in treatments develops differently in the remaining stages. Although possibly endogenous,[36] females significantly worsen their already low performance in the RPF-ind-ind ($p < 0.10$) relative to APF-ind-ind. Moreover females in RPF-team-team improves ($p < 0.05$), and RPF-team-ind ($p < 0.01$) worsen their performance, relative to APF-team-team. The development of females in RPF-ind-ind is significantly negative relative to the development of males in RPF-ind-ind ($p < 0.01$),

[36] We use the first stage as a control, and this may be correlated with both their performance and the treatment they are in.

Table 5
Quantile regressions of treatments.

Quantile:	10% (1)	25% (2)	50% (3)	75% (4)	90% (5)
APF-team-team	Ref.	Ref.	Ref.	Ref.	Ref.
APF-ind-ind	1.396	2.419	2.220	3.331**	1.538
	(1.6957)	(1.4734)	(1.7520)	(1.3587)	(1.9751)
RPF-ind-ind	1.104	2.510	2.076	4.369**	5.318**
	(2.2584)	(1.9178)	(1.4393)	(2.0998)	(2.6197)
RPF-ind-team	−0.021	2.763	3.890*	5.755***	6.674***
	(1.8944)	(1.8080)	(2.1283)	(2.0922)	(2.2042)
RPF-team-team	−0.417	0.545	1.227	5.866***	7.242***
	(1.6717)	(1.6107)	(2.0427)	(2.0212)	(2.4750)
RPF-team-ind	−0.479	1.035	1.892	1.477	2.674
	(2.3413)	(1.8732)	(1.6899)	(1.9261)	(2.2199)
Constant	32.896***	37.541***	36.911***	44.503***	49.333***
	(3.8628)	(4.4789)	(3.7987)	(4.3926)	(3.9050)
Observations	338	338	338	338	338

Notes: Quantile regression coefficients reported, with robust standard errors in parentheses, based on bootstrapping with 1.000 replications. Dependent variable is the average number of solved tasks across all stages. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table 6
Gender analysis.

	All stages (1)	Stages 2–6 (2)
APF-team-team	Ref.	Ref.
APF-ind-ind	2.327***	−0.253
	(0.8962)	(0.7146)
RPF-ind-ind	4.694***	0.243
	(0.6159)	(0.9113)
RPF-ind-team	3.276***	0.113
	(1.1183)	(0.7562)
RPF-team-team	3.525***	0.614
	(1.2548)	(0.6985)
RPF-team-ind	1.491***	−1.105
	(0.4699)	(0.7579)
Female	−1.200	0.276
	(1.3425)	(0.5205)
APF-ind-ind x Female	0.706	−0.725
	(2.0603)	(0.9747)
RPF-ind-ind x Female	−2.680*	−2.161***
	(1.4489)	(0.7401)
RPF-ind-team x Female	2.360	−0.066
	(4.0590)	(1.5876)
RPF-team-team x Female	−0.102	1.315
	(1.7136)	(1.2647)
RPF-team-ind x Female	0.589	−1.217
	(1.3899)	(0.8489)
Stage $_t$	2.384***	2.031***
	(0.0773)	(0.0797)
Number of clusters	18	18
Observations	2028	1690

Notes: Random Effects GLS coefficients reported, with robust standard errors in parentheses, corrected for clustering across sessions. The dependent variable in column (1) is number of solved tasks in all stages, whereas in column (2) it is number of solved tasks in stages 2–6, only with a control for the 1st stage. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for economics students and a dummy for Norwegian nationality. Constant and 1st stage variable is also omitted from the table. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

further strengthening the gender difference after the first stage of this treatment.

Further exploring gender differences; consider now the interaction effects between feedback and incentives. In Table 7, we employ the same analysis as in Section 4.3, but on each gender separately. There are two apparent gender differences to notice. First, males respond strongly negative to team incentives, whereas females do not. Second, there is a strongly positive effect of team RPF among females, which for males goes the opposite direction. Hence, while males are triggered by individual RPF, it is team RPF that triggers females. The positive complementarity between team incentives and team RPF amongst males, however, offset the negative effects (i.e. the sum of the coefficients team incentives, team RPF and the interaction between them is not significantly different from zero). Females, on the other hand, only need team RPF to improve performance, and do not gain additional productivity when interacting the two variables. Despite this, females performance with both team incentives and team RPF is greater than individual incentives and individual RPF ($p < 0.05$). This is consistent with H4, suggesting that females put more weight on social comparison (θ) in team-based environments, and that team RPF can lead females to improve their effort and performance.

Result 5: *We find suggestive evidence that females respond positively to team RPF relative to other treatments, and evidence that they do not respond to individual RPF. This supports H4.*

4.5. Results from the team leader treatments

We have shown that when subjects are exposed to team incentives, then team RPF increases the team's average performance significantly.

Table 7
Changing incentives and feedback – gender analysis.

Panel:	Males (1)	Females (2)
Individual incentives and individual RPF	Ref.	Ref.
Team incentives	−3.398***	−0.873
	(1.1922)	(1.0125)
Team RPF	−2.171**	3.816**
	(1.1002)	(1.6738)
Team incentives X Team RPF	2.208	0.554
	(2.2417)	(2.4686)
Stage $_t$	2.377***	2.259***
	(0.1225)	(0.1945)
Constant	38.734***	29.405***
	(5.0498)	(3.1891)
Number of clusters	12	12
Observations	732	546

Notes: Random Effects GLS coefficients reported, with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks across all stages. Both columns include the following control variables: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

We have also shown that team incentives and team RPF are complements. It remains to identify the mechanism behind these results. In the theoretical framework, we present two potential mechanisms: Peer pressure /team spirit and competitive preferences/status concerns. The strong effect we find on top performers and the insignificant effect on other team members indicate that team spirit may not be the main explanation to our results. Our team leader treatments are meant to further explore this. The approach is to reduce peer pressure, at least in terms of peer pressure from free-riding and underperforming within the team, by letting people *work on behalf of teams as team leaders*, where the others in the team do not contribute to the team output.

The results are as follows: Average performance is significantly greater for subjects in RPF-teamleader than in APF-teamleader in both stage 1 (MW: $p = 0.085$, RT: $p = 0.050$) and stage 2 (MW: $p = 0.082$, RT: $p = 0.037$). The average difference across all stages is not statistically significant (MW: $p = 0.249$, RT= $p = 0.162$), but the gap in number of solved tasks remains more than one task throughout all six stages. In Table 8, we run regressions and find the effect in the 2nd stage to be significant at the 5% level. Fig. A3 displays the development in performance across stages for both treatments. Gender analysis show that females are the main driver behind the difference, but there are no statistical gender difference in performance.

One should be careful comparing the two team leader treatments with the previous treatments since they were not run at the same time. However, it is worth noting that performance under RPF-teamleader and RPE-team-team are almost the same. Hence, reducing peer pressure when subjects are exposed to team RPF does not affect performance. It is also worth noting that subjects in APF-teamleader do significantly better than subjects in APF-team-team, suggesting that the team leader *framing* may in itself be motivating.

Result 6: *Team leaders who receive team RPF outperform team leaders who receive team APF.*

This result, together with the top performer result (Result 4), indicates that the main driver behind the effects of team RPF is status concerns or competitive preferences, and not team spirit. However, our experimental results cannot rule out that team spirit also contribute to the positive effect of team RPF. In particular, one should beware that the positive effects of RPF is lower in the team leader treatments. Furthermore, there is also the possibility that status concerns are even more prominent in the team leader treatments, i.e. status concerns are not

Table 8
Team leader results: effects on productivity.

Stage(s):	1st stage (1)	2nd stage (2)	All stages (3)
APF-team-leader	Ref.	Ref.	Ref.
RPF-team-leader	1.383*	1.461**	0.840
	(0.6693)	(0.6391)	(0.7364)
Stage $_t$			2.473***
			(0.0674)
Constant	27.728***	32.078***	31.996***
	(2.3840)	(2.4298)	(2.7869)
Adjusted R^2	0.058	0.061	
Observations	177	177	1062

Notes: OLS coefficients reported in columns (1) – (2) and Random Effects GLS coefficients reported in column (3), with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

held constant when moving to the team leader treatments. If that is the case, it might be that a reduction in performance due to lower peer pressure is balanced out by higher performance due to stronger social concerns. Hence, more research is needed to fully understand the mechanisms behind our results.

5. Concluding remarks

In this paper, we investigate experimentally how teams respond to relative performance feedback (RPF). We find that when subjects are exposed to team incentives, then RPF on how their team is doing compared to two other teams increases the team's average performance by almost 10%. The treatment effect is driven by the teams' top performers. The average individual performance of the top performers within each team is almost 20% higher when the teams receive relative performance feedback compared to when the teams only receive absolute performance feedback. Our experiment suggests that subjects, and in particular top performers, are motivated by the combination of team incentives and team RPF.

This result complements the interesting and somehow puzzling findings by Hamilton et al. (2003), namely that high ability workers were more attracted to team work than low ability workers. When offering workers at a garment plant the opportunity to shift from individual piece rates to team incentives, the high-productivity workers tended to join teams first, despite a loss in earnings for many of them. Hamilton et al. (2003) suggested that high-ability workers may acquire a higher social status in teams and are therefore willing to join teams even if their own pay is reduced. Our results illuminate their findings, which suggest that high ability workers are not motivated by team incentives alone. Rather, they seem to be motivated by the chance to help the team achieve some non-monetary goals, which in our experiment is higher ranking.

Our results from the team leader treatments further support this conjecture. In the team leader treatments, we removed (or at least reduced) peer pressure by letting people work on behalf of teams, where the others in the team did not work. We find that team leaders receiving RPF perform significantly better than team leaders who only receive absolute performance feedback, indicating that status concerns or competitive preferences better explain our results than peer pressure or team spirit.

For managers designing feedback interventions in their organization, there are several implications of this experiment. First, competition between teams for higher ranks may be an efficient way to improve the productivity of employees, in particular if they are paid as a team. Second, teamwork does not suppress top performance. On the contrary, team competition may be an efficient way of motivating high ability workers. Third, team feedback is a good alternative to individual feedback in organizations with significant shares of female workers. Females, who are more negatively inclined to individual RPF, seem to be particularly productive when they are provided with team performance data rather than individual performance data.

Acknowledgements

We would like to thank the editor, three anonymous reviewers, Robert Dur, Uri Gneezy, Nagore Iriberri, Petra Nieken, Mari Rege, Dirk Sliwka, and seminar participants at the Rady School of Management, University of Cologne, University of Oslo, Norwegian School of Economics and Moscow Higher School of Economics for helpful comments and suggestions. Financial support from the Norwegian Research Council (227004) is gratefully acknowledged.

Appendix

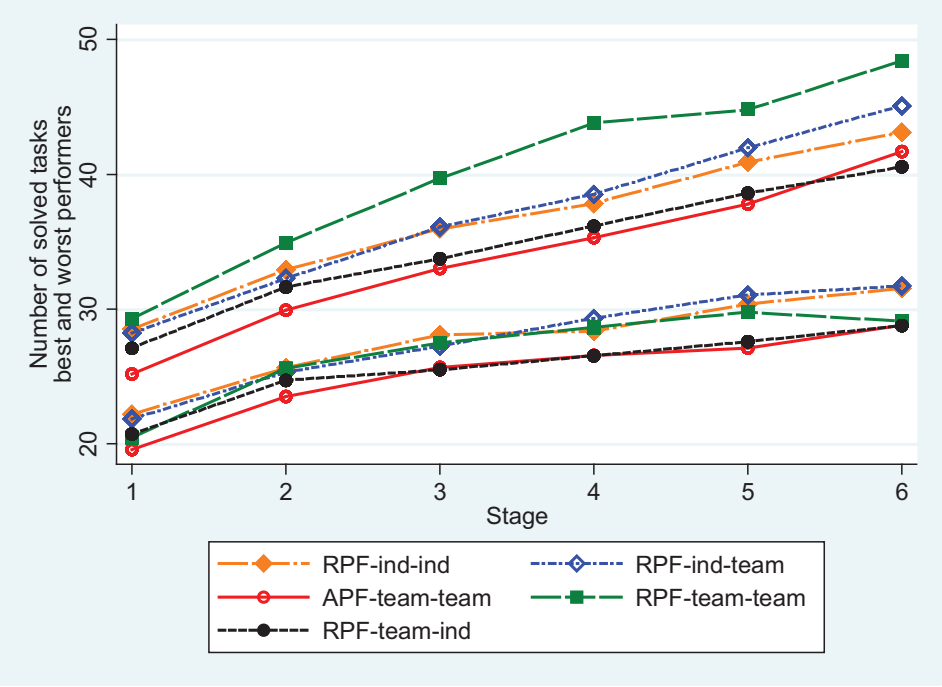

Fig. A1. Best and worst performers across treatments
Note: best and worst performers are categorized based on their performance in the final working stage. See the start of Section 4.4 for further details.

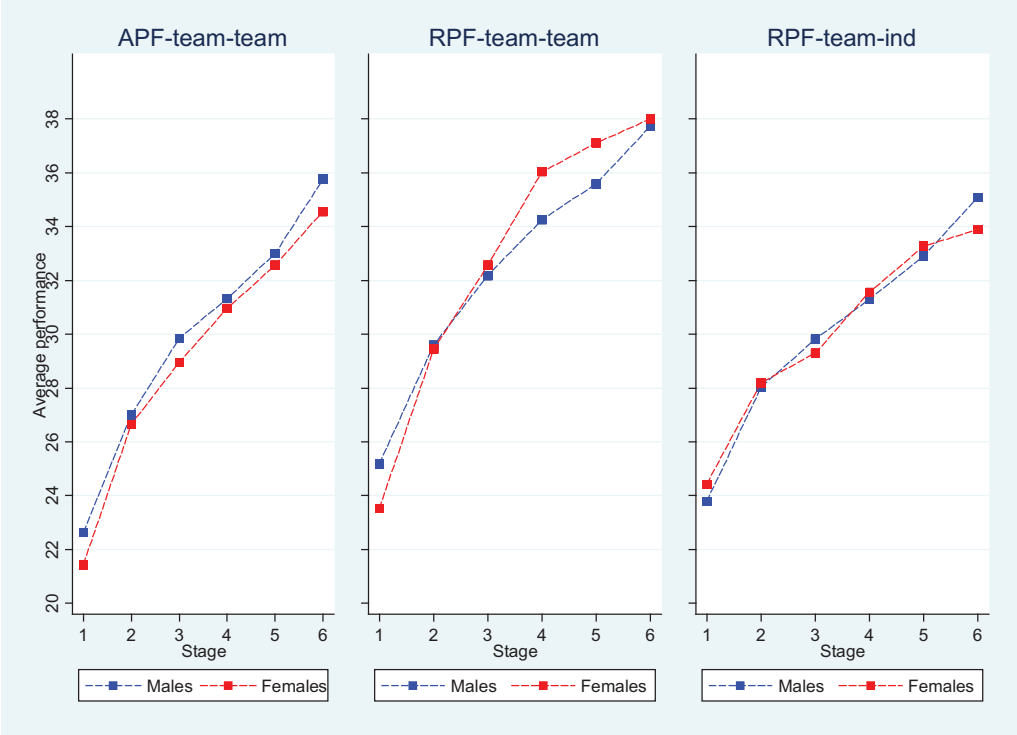

Fig. A2. Gender – team incentive treatments.

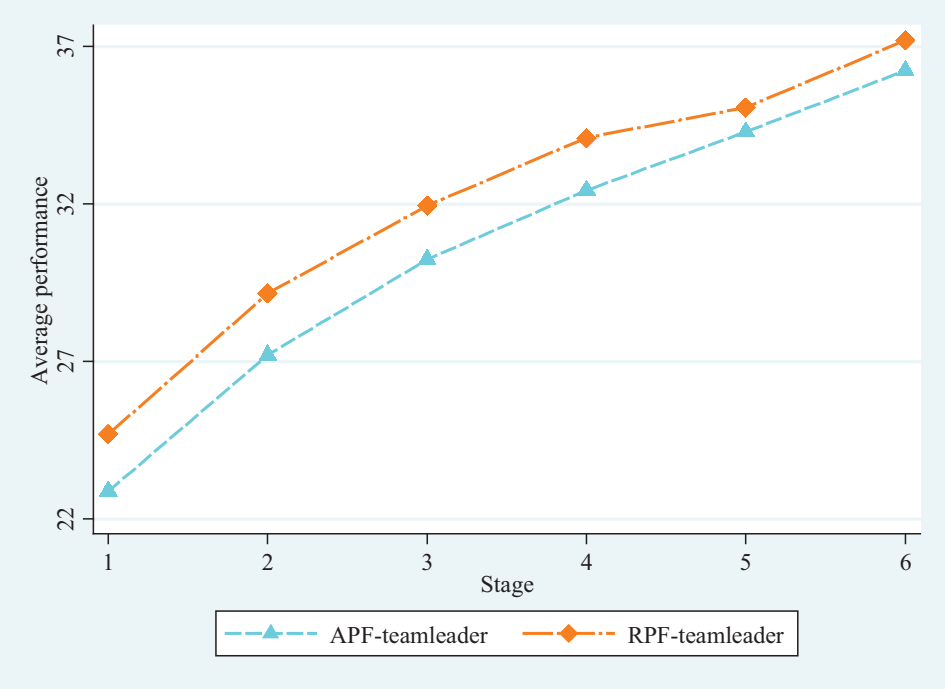

Fig. A3. Average performance across stages in teamleader treatments.

Table A1
Summary statistics of control variables.

	APF-ind-ind (1)	RPF-ind-ind (2)	RPF-ind-team (3)	APF-team-team (4)	RPF-team-team (5)	RPF-team-ind (6)	Pearson^2/ Kruskal Wallis (7)	APF-team-leader (8)	RPF-team-leader (9)	Pearson^2/Kruskal Wallis (10)
Economics-students	0.132	0.039	0.044	0.123	0.204	0.143	0.237	0.226	0.083	0.009
	(0.341)	(0.196)	(0.208)	(0.331)	(0.407)	(0.353)		(0.420)	(0.278)	
Norwegian-Nationality	0.706	0.510	0.444	0.579	0.519	0.413	0.008	0.559	0.560	0.996
	(0.459)	(0.505)	(0.503)	(0.498)	(0.504)	(0.496)		(0.499)	(0.499)	
Age	24.29	26	26.31	24.25	25.57	25.35	0.025	26.04	25.37	0.879
	(4.316)	(4.060)	(5.049)	(3.291)	(4.364)	(4.656)		(7.228)	(4.935)	
Female	0.426	0.392	0.311	0.404	0.500	0.476	0.365	0.538	0.583	0.541
	(0.498)	(0.493)	(0.468)	(0.495)	(0.505)	(0.503)		(0.501)	(0.496)	
Average-grade	2.559	2.078	2.311	2.526	2.370	2.508	0.008	2.559	2.500	0.368
	(0.720)	(0.796)	(0.668)	(0.782)	(0.623)	(0.592)		(0.787)	(0.768)	
Observations	68	51	45	57	54	63	338	93	84	177

Notes: Mean and (standard deviation). For columns (1) to (6) we report p-value of Pearson^2 for binary variables and Kruskal Wallis for non-binary variables in column (7). For columns (8) to (9) we report p-value of Pearson^2 for binary variables and Kruskal Wallis for non-binary variables in column (10).

Table A2
Treatment effects across stages.

Stages:	3rd stage (1)	4th stage (2)	5th stage (3)	6th stage (4)
APF-team-team	Ref.	Ref.	Ref.	Ref.
APF-ind-ind	2.235***	2.652***	2.520***	2.757***
	(0.7610)	(0.6004)	(0.5537)	(0.8270)
RPF-ind-ind	3.233***	3.299***	4.015***	3.189**
	(0.9169)	(0.6593)	(0.9653)	(1.1490)
RPF-ind-team	3.058***	4.280***	4.834***	4.261***
	(0.7963)	(0.6435)	(0.6845)	(0.8472)
RPF-team-team	3.406**	4.373***	4.183***	3.323*
	(1.2906)	(1.1914)	(1.3047)	(1.6776)
RPF-team-ind	0.695	1.472**	1.617**	0.389
	(0.8528)	(0.5808)	(0.6816)	(0.9743)
Constant	39.939***	44.014***	45.520***	48.121***
	(2.6572)	(2.5693)	(3.1176)	(3.5823)
Adjusted R^2	0.095	0.132	0.117	0.086
Number of clusters	18	18	18	18
Observations	338	338	338	338

Notes: OLS coefficients reported, with robust standard errors in parenthesis, corrected for clustering across sessions. Dependent variable is number of solved tasks. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table A3
Mann-Whitney pairwise test: Performance across incentive scheme.

	Average Performance (SD)		Mann-Whitney z-Statistics
	APF-ind-ind (1)	APF-team-team (2)	(*p*-value) (1) vs. (2)
Stage 1	25.31 (4.90)	22.16 (5.77)	3.16 (0.002)***
Stage 2	28.97 (5.32)	26.86 (4.50)	2.34 (0.020)**
All stages	32.43 (6.06)	29.63 (5.56)	2.57 (0.010)**
N	68	57	125

Notes: * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table A4
Mann-Whitney pairwise test: Team incentives and RPF.

	Average Performance (SD)			Mann-Whitney z-Statistics	
	APF-team-team (1)	RPF-team-team (2)	RPF-team-ind (3)	(*p*-value) (1) vs. (2)	(1) vs. (3)
Stage 1	22.16 (5.77)	24.35 (6.81)	24.10 (4.76)	−1.48 (0.138)	−1.36 (0.175)
Stage 2	26.86 (4.49)	29.52 (6.23)	28.11 (5.06)	−2.06 (0.040)**	−1.22 (0.223)
All stages	29.63 (5.56)	32.60 (7.63)	30.13 (5.51)	−1.69 (0.090)*	−0.21 (0.834)
N	57	54	63	111	120

Notes: * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Table A5
Persistence of treatment effects.

Stages:	Stages 1–3 (1)	Stages 4–6 (2)	Stages 2–6 (3)
APF-team-team	Ref.	Ref.	Ref.
APF-ind-ind	2.465***	2.699***	2.457***
	(0.6269)	(0.5486)	(0.5321)
RPF-ind-ind	3.630***	3.560***	3.437***
	(0.9090)	(0.8195)	(0.7588)
RPF-ind-team	3.371***	4.462***	3.877***
	(0.7934)	(0.6217)	(0.5986)
RPF-team-team	3.032**	4.018***	3.719***
	(1.3435)	(1.2711)	(1.2176)
RPF-team-ind	1.945**	1.329**	1.314**
	(0.7668)	(0.6507)	(0.5885)
Stage $_t$	3.379***	1.880***	2.031***
	(0.1372)	(0.1135)	(0.0796)
Constant	28.678***	36.947***	34.669***
	(2.4569)	(3.1182)	(2.7724)
Number of clusters	18	18	18
Observations	1014	1014	1690

Notes: Random Effects GLS coefficients reported, with robust standard errors in parentheses, corrected for clustering across sessions. Dependent variable is number of solved tasks. All columns have the following control variables included: Time on the day of the session (FE in panel), age, average grades at University level, a dummy for gender, a dummy for economics students and a dummy for Norwegian nationality. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6
Mann-Whitney pairwise test: Individual incentives and RPF.

	Average Performance (SD)			Mann-Whitney z-Statistics	
	APF-ind-ind (1)	RPF-ind-ind (2)	RPF-ind-team (3)	(p-value) (1) vs. (2)	(1) vs. (3)
Stage 1	25.31 (4.90)	25.49 (6.06)	25.13 (5.02)	0.34 (0.737)	0.34 (0.735)
Stage 2	28.97 (5.32)	29.53 (6.11)	29.20 (5.52)	−0.07 (0.949)	−0.27 (0.787)
All stages	32.43 (6.06)	32.29 (7.11)	32.46 (6.08)	0.81 (0.421)	0.22 (0.828)
N	68	51	45	119	113

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7
Mann-Whitney pairwise test: Team RPF compared to individual incentives.

	Average Performance (SD)			Mann-Whitney z-Statistics	
	APF-ind-ind (1)	RPF-ind-ind (2)	RPF-team-team (3)	(*p*-value) (1) vs. (3)	(2) vs. (3)
Stage 1	25.31 (4.90)	25.49 (6.06)	24.35 (6.81)	1.20 (0.230)	0.81 (0.420)
Stage 2	28.97 (5.32)	29.53 (6.11)	29.52 (6.23)	0.01 (0.994)	0.10 (0.923)
All stages	32.43 (6.06)	32.29 (7.11)	32.60 (7.63)	0.46 (0.648)	−0.09 (0.926)
N	68	51	54	122	105

Notes: * $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

Experimental Instructions

Welcome to the experiment (APF- ind - ind)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, you will earn 1 NOK for each task you solve. In other words, your payment depends on how many tasks you solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (RPF- ind - ind)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, you will earn 1 NOK for each task you solve. In other words, your payment depends on how many tasks you solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

In addition, your performance will be ranked relative to two other randomly selected participants in the room, and you will be informed about how many tasks they have solved. You will be ranked relative to the same participants in all of the breaks. Ranks will not affect your payment.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (RPF- ind -team)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room, and you will all be working simultaneously on the same type of tasks. The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, you will earn 1 NOK for each task you solve. In other words, your payment depends on how many tasks you solve. Your payment does not depend on how many tasks the other team members solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

In addition, you will also be informed about the total output of your team in the previous stage. Also, your team performance will be ranked relative to two other teams in the room, and you will be informed about how many tasks these teams have solved. Your team will be ranked relative to the same teams in all of the breaks. Ranks will not affect your payment.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (APF-team-team)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room, and you will all be working simultaneously on the same type of tasks. The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, your team will earn 1 NOK for each task a team member solves. The total earnings of the team is then divided equally among each team member independently of actual contribution. In other words, your payment depends on how many tasks you and your other team members solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

In addition, you will also be informed about the total output of your team in the previous stage.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (RPF-team-team)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room, and you will all be working simultaneously on the same type of tasks. The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, your team will earn 1 NOK for each task a team member solves. The total earnings of the team is then divided equally among each team member independently of actual contribution. In other words, your payment depends on how many tasks you and your other team members solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

In addition, you will also be informed about the total output of your team in the previous stage. Also, your team performance will be ranked relative to two other teams in the room, and you will be informed about how many tasks these teams have solved. Your team will be ranked relative to the same teams in all of the breaks. Ranks will not affect your payment.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (RPF-team- ind)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room, and you will all be working simultaneously on the same type of tasks. The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, your team will earn 1 NOK for each task a team member solves. The total earnings of the team is then divided equally among each team member independently of actual contribution. In other words, your payment depends on how many tasks you and your other team members solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you have correctly solved and how much you have earned during the previous stage.

In addition, you will also be informed about the total output of your team in the previous stage. Also, your contribution to the team performance will be ranked relative to the other two team members, and you will be informed about how many tasks each team member have solved. Ranks will not affect your payment.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (APF teamleader)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room. **You are selected as the team leader.** The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, your team will earn 1 NOK for each task you as the team leader solves. The total earnings is then divided equally among each team member. In other words, your payment (as well as the team's payment) depends on how many tasks you as the team leader solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you as the team leader have correctly solved on behalf of the team and how much you and your team have earned during the previous stage.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

Welcome to the experiment (RPF teamleader)

Task description:

We ask you to decode letters into numbers. You are given a list of letters, all of which have been assigned with a corresponding number. Your task is then to decode given sequences of four letters into numbers.

Example: Given this list of letters

A	B	C	D	E	F	G
8	12	14	10	9	6	24

Task-

Decode these letters: **A | E | G | F**

Correct answer: **8 | 9 | 24 | 6**

Stages and process of the experiment:

The experiment consists of six working stages, and the duration of each stage is five minutes. There is an unlimited number of tasks in each stage. A countdown in the upper right corner of the computer screen displays remaining time of current stage. After the final stage, we will ask you to fill out a short questionnaire. Total duration of the experiment is estimated to be about 45 min.

Team:

You are part of a team consisting of a total of three randomly selected participants in the room. **You are selected as the team leader.** The team will remain unchanged throughout the experiment.

Payment:

Everyone earns 100 NOK for participating in the experiment. In addition, your team will earn 1 NOK for each task you as the team leader solves. The total earnings is then divided equally among each team member. In other words, your payment (as well as the team's payment) depends on how many tasks you as the team leader solve.

Breaks:

In between each stage there will be a minute break. During the breaks, you will be provided with information about how many tasks you as the team leader have correctly solved on behalf of the team and how much you and your team have earned during the previous stage.

In addition, your performance as team leader will be ranked relative to two other team leaders in the room, and you will be informed about how many tasks these team leaders have solved. Your team will be ranked relative to the same teams in all of the breaks. Ranks will not affect your payment.

Rules:

You choose freely how to spend your time during the experiment. However, we do require you to remain in your seat throughout the experiment, and refrain from communicating with other participants. You may use your mobile phone to surf the internet, but please ensure that it is in a mute state before we start. It is strictly prohibited to use the pc for anything other than the experiment, as different usage may cause technical problems with the experiment.

Thank you for participating in the experiment.

References

Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Q. J. Econ. 115 (3), 715–753.

Akerlof, G.A., Kranton, R.E., 2005. Identity and the economics of organizations. J. Econ. Perspect. 19 (1), 9–32.

Alchian, A.A., Demsetz, H., 1972. Production, information costs, and economic organization. Am. Econ. Rev. 62 (5), 777–795.

Auriol, E., Renault, R., 2008. Status and incentives. Rand. J. Econ. 39 (1), 305–326.

Azmat, G., Iriberri, N., 2016. The provision of relative performance feedback: an analysis of performance and satisfaction. J. Econ. Manag. Strat. 25 (1), 77–110.

Babcock, P., Bedard, K., Charness, G., Hartman, J., Royer, H., 2015. Letting down the team? Social effects of team incentives. J. Eur. Econ. Assoc. 13 (5), 841–870. doi:10.1111/jeea.12131.

Baer, M., Leenders, R.T.A., Oldham, G.R., Vadera, A.K., 2010. Win or lose the battle for creativity: the power and perils of intergroup competition. Acad. Manag. J. 53 (4), 827–845.

Bandiera, O., Barankay, I., Rasul, I., 2013. Team incentives: evidence from a firm level experiment. J. Eur. Econ. Assoc. 11 (5), 1079–1114.

Barankay, I., 2012. Rank Incentives: Evidence From a Randomized Workplace Experiment. The Wharton School Working paper.

Bellemare, C., Lepage, P., Shearer, B., 2010. Peer pressure, incentives, and gender: an experimental analysis of motivation in the workplace. Labour Econ. 17 (1), 276–283.

Bertrand, M., 2011. In: Card, D., Ashenfelter, O. (Eds.). In: Chapter 17 - New perspectives on Gender, 4. Elsevier, pp. 1543–1590 Part B.

Birkinshaw, J., 2001. Why is knowledge management so difficult? Bus. Strat. Rev. 12 (1), 11–18.

Vidal, J., Nossol, M., 2011. Tournaments without prizes: evidence from personnel records. Manag. Sci. 57 (10), 1721–1736.

Boning, B., Ichniowski, C., Shaw, K., 2007. Opportunity counts: teams and the effectiveness of production incentives. J. Labor Econ. 25 (4), 613–650.

Bornstein, G., Ben-Yossef, M., 1994. Cooperation in intergroup and single-group social dilemmas. J. Exp. Soc. Psychol. 30 (1), 52–67.

Bornstein, G., Erev, I., Rosen, O., 1990. Intergroup competition as a structural solution to social dilemmas. Soc. Behav. 5 (4), 247–260.

Bornstein, G., Gneezy, U., Nagel, R., 2002. The effect of intergroup competition on group coordination: an experimental study. Games Econ. Behav. 41 (1), 1–25.

Burton-Chellew, M.N., West, S.A., 2012. Pseudocompetition among groups increases human cooperation in a public-goods game. Anim. Behav. 84 (4), 947–952.

Büyükboyaci, M., Robbett, A., 2019. Team formation with complementary skills. J. Econ. Manag. Strat. 28 (4), 713–733. doi:10.1111/jems.12296.

Büyükboyacı, M., Robbett, A., 2017. Collaboration and free-riding in team contests. Labour Econ. 49, 162–178.

Böhm, R., Rockenbach, B., 2013. The inter-group comparison–intra-group cooperation hypothesis: comparisons between groups increase efficiency in public goods provision. PLoS ONE 8 (2), e56152.

Charness, G., Gneezy, U., 2012. Strong evidence for gender differences in risk taking. J. Econ. Behav. Organ. 83 (1), 50–58.

Charness, G., Grosskopf, B., 2001. Relative payoffs and happiness: an experimental study. J. Econ. Behav. Organ. 45 (3), 301–328.

Charness, G., Masclet, D., Villeval, M.C., 2014. The dark side of competition for status. Manag. Sci. 60 (1), 38–55.

Che, Y.-.K., Seung-Weon, Y., 2001. Optimal incentives for teams. Am. Econ. Rev. 91 (3), 525–541.

Chen, H., Lim, N., 2013. Should managers use team-based contests? Manag. Sci. 59 (12), 2823–2836. doi:10.1287/mnsc.2013.1743.

Chen, R., Gong, J., 2018. Can self selection create high-performing teams? J. Econ. Behav. Organ. 148, 20–33.

Clark, A.E., Oswald, A.J., 1996. Satisfaction and comparison income. J. Public Econ. 61 (3), 359–381.

Corgnet, B., Hernán-González, R., Rassenti, S., 2015. Peer pressure and moral hazard in teams: experimental evidence. Rev. Behav. Econ. 2 (4), 379–403.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (2).

Dargnies, M.-.P., 2012. Men too sometimes shy away from competition: the case of team competition. Manag. Sci. 58 (11), 1982–2000.

Delfgaauw, J., Dur, R., Sol, J., Verbeke, W., 2013. Tournament incentives in the field: gender differences in the workplace. J. Labor Econ. 31 (2), 305–326.

Erev, I., Bornstein, G., Galili, R., 1993. Constructive intergroup competition as a solution to the free rider problem: a field experiment. J. Exp. Soc. Psychol. 29 (6), 463–478.

Eriksson, T., Poulsen, A., Villeval, M.C., 2009. Feedback and incentives: experimental evidence. Labour Econ. 16 (6), 679–688.

Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.

Flory, J.A., Leibbrandt, A., List, J.A., 2015. Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions. Rev. Econ. Stud. 82 (1), 122–155.

Frank, R.H., 1985. Choosing The Right Pond: Human Behavior and The Quest For Status. Oxford University Press.

Gjedrem, W.G., 2018. Relative performance feedback: effective or dismaying? J. Behav. Exp. Econ. 74, 1–16.

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. Q. J. Econ. 118 (3), 1049–1074.

Gneezy, U., Rustichini, A., 2004. Gender and competition at a young age. Am. Econ. Rev. 94 (2), 377–381.

Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer effects in the workplace: evidence from random groupings in professional golf tournaments. Am. Econ. J. Appl. Econ. 1 (4), 34–68.

Hamilton, B.H., Nickerson, J.A., Owan, H., 2003. Team incentives and worker heterogeneity: an empirical analysis of the impact of teams on productivity and participation. J. Polit. Econ. 111 (3), 465–497.

Hannan, R.L., Krishnan, R., Newman, A.H., 2008. The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. Account. Rev. 83 (4), 893–913.

Hannan, R.L., McPhee, G.P., Newman, A.H., Tafkov, I.D., 2013. The effect of relative performance information on performance and effort allocation in a multi-task environment. Account. Rev. 88 (2), 553–575.

Healy, A., Pate, J., 2011. Can teams help to close the gender competition gap? Econ. J. 121 (555), 1192–1204.

Herbst, L., Konrad, K.A., Morath, F., 2015. Endogenous group formation in experimental contests. Eur. Econ. Rev. 74, 163–189.

Holmstrom, B., 1982. Moral hazard in teams. Bell J. Econ. 324–340.

Holmström, B., Milgrom, P., 1990. Regulating trade among agents. J. Inst. Theor. Econ. 85–105.

Imbens, G.W., Rubin, D.B., 2015. Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press.

Itoh, H., 1991. Incentives to help in multi-agent situations. Econometrica 59 (3), 611–636.

Itoh, H., 1992. Cooperation in hierarchical organizations: an incentive perspective. J. Law Econ. Organ. 8 (2), 321–345.

Kaiser, J., 2007. An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples. Stata J. 7 (3), 402–412.

Kandel, E., Lazear, E.P., 1992. Peer pressure and partnerships. J. Polit. Econ. 100 (4), 801–817.

Knez, M., Simester, D., 2001. Firm-wide incentives and mutual monitoring at Continental Airlines. J. Labor Econ. 19 (4), 743–772.

Kramer, S., Maas, V.S., Rinsum, M.v., 2016. Relative performance information, rank ordering and employee performance: a research note. Manag. Account. Res. 33, 16–24.

Kuhn, P., Villeval, M.C., 2015. Are women more attracted to co-operation than men? Econ. J. 125 (582), 115–140.

Kuhnen, C.M., Tymula, A., 2012. Feedback, self-esteem, and performance in organizations. Manag. Sci. 58 (1), 94–113.

Kvaløy, O., Olsen, T.E., 2006. Team incentives in relational employment contracts. J. Labor Econ. 24 (1), 139–169.

Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89 (5), 841–864.

Lazear, E.P., Shaw, K.L., 2007. Personnel economics: the economist's view of human resources. J. Econ. Perspect. 21 (4), 91–114.

Macho-Stadler, I., Pérez-Castrillo, J.D., 1993. Moral hazard with several agents: the gains from cooperation. Int. J. Ind. Organ. 11 (1), 73–100.

Marino, A.M., Zábojnik, J., 2004. Internal Competition for Corporate Resources and Incentives in Teams. Rand J. Econ. 35 (4), 710–727.

Nalbantian, H.R., Schotter, A., 1997. Productivity under group incentives: an experimental study. Am. Econ. Rev. 87 (3), 314–341.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much. Q. J. Econ. 122 (3), 1067–1101.

Sausgruber, R., 2009. A note on peer effects between teams. Exp. Econ. 12 (2), 193–201.

Sheremeta, R.M., 2018. Behavior in group contests: a review of experimental research. J. Econ. Surv. 32 (3), 683–704.

Tafkov, I.D., 2013. Private and public relative performance information under different compensation contracts. Account. Rev. 88 (1), 327–350.

Tan, J.H., Bolle, F., 2007. Team competition and the public goods game. Econ. Lett. 96 (1), 133–139.

Turner, J.C., 1975. Social comparison and social identity: Some prospects for intergroup behaviour. Eur. J. Soc. Psychol. 5 (1), 1–34.

van Dijk, F., Sonnemans, J., van Winden, F., 2001. Incentive systems in a real effort experiment. Eur. Econ. Rev. 45 (2), 187–214.

Vandegrift, D., Yavas, A., 2011. An experimental test of behavior under team production. Manag. Decis. Econ. 32 (1), 35–51.

Young, A., 2018. Channeling Fisher: randomization tests and the statistical insignificance of seemingly significant experimental results*. Q. J. Econ. 134 (2), 557–598.