CESifo WORKING PAPERS

6871
2018

January 2018

Relative Performance Feedback to Teams

William Gilje Gjedrem, Ola Kvaløy

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6871*
Category 13: Behavioural Economics

Relative Performance Feedback to Teams

Abstract

Between and within firms, work teams compete against each other and 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) at team level. We find that when subjects work under team incentives, then RPF on team performance increases the teams' average performance by almost 10 percent. 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. In fact, team incentives motivate significantly higher top performance than individual incentives when the team is exposed to RPF. We also find notable gender differences. Females respond negatively to individual RPF, but even more positively than males to team RPF.

JEL-Codes: C910, M500, M520.

Keywords: teams, performance feedback, performance pay, experiment.

William Gilje Gjedrem
University of Stavanger
Norway – 4036 Stavanger
william.g.gjedrem@uis.no

Ola Kvaløy
University of Stavanger
Norway – 4036 Stavanger
ola.kvaloy@uis.no

We thank Petra Nieken, Mari Rege and seminar participants at the Stavanger Workshop on Incentives and Motivation, Rady School of Management, University of Cologne, University of Oslo, ESA meeting in Bergen, and the Choice Lab at the Norwegian School of Economics for helpful comments and suggestions. Financial support from the Norwegian Research Council (227004) is gratefully acknowledged.

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 economists have systematically studied how people respond to RPF. The early economics literature on relative performance evaluation studied the effect of connecting rank to monetary incentives (see Lazear and Rosen (1981)'s seminal contribution on rank order tournaments). More recent theories on competitive preferences and status concerns (Frank, 1985; Clark & Oswald, 1996; Auriol & Renault, 2008) suggest, however, that rank *per se* motivates effort.[1] And 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) and Charness, Masclet, and Villeval (2014), find strong performance improvements in situations where RPF is provided, while Hannan, Krishnan, and Newman (2008), Gjedrem (2015), and Azmat and Iriberri (2016) find significant context specific effects of RPF. There are also studies that do not find any positive effects of RPF. Guryan, Kroft, and Notowidigdo (2009), Eriksson, Poulsen, and Villeval (2009) and Bellemare, Lepage, and Shearer (2010) find no significant effects, while Barankay (2012) find that removing RPF positively affected productivity.

Except for a field experiment by Delfgaauw, Dur, Sol, and Verbeke (2013), the experimental literature on RPF has so far concentrated on individual behavior and individual feedback.[2] However, not only individuals receive RPF, but also 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 & Zábojnik, 2004; Baer, Leenders, Oldham, & Vadera, 2010). Successful teams are typically compensated by some monetary rewards, but team competitions *per se* may also be motivating.

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

[2] Delfgaauw et al. (2013) studies 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.

We contribute to the existing literature by investigating how teams respond to relative performance feedback. From a theory perspective, there are mainly 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 be different from the utility of winning alone. Similarly, the costs of losing as a team might be different from the costs of losing alone. The second relates to peer pressure and "team spirit". As demonstrated theoretically (Kandel & Lazear, 1992) and empirically (e.g., Babcock, Bedard, Charness, Hartman, & Royer, 2015; Corgnet, Hernán-González, & Rassenti, 2015), peer pressure can motivate workers in teams to exert effort. 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.

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.

Our main 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 percent. The treatment effect is driven by higher top performances. The average individual performance of the best performance within each team is almost 20 % higher when the teams receive RPF compared to when the teams only receive APF. These effects more or less disappear under individual incentives and/or individual RPF. Our experiment thus suggests that some subjects are particularly motivated by the combination of team incentives and team RPF. In fact, team incentives motivate significantly higher top performance than individual incentives, when subjects are exposed to team RPF. The strong effect on top performers, and the insignificant effect on other team members

indicate that team spirit is not a main explanation of our results. Our results from the team leader treatments supports 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 complements 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 free-rider problems. However, our results contrast with a recent field study by Bandiera et al. (2013). They find that ranking teams reduces overall performance, as lower ranked teams decrease productivity. But they endogenously allows subjects to select into teams, whereas we assign subjects exogenously into teams.

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 & Gneezy, 2009; Bertrand, 2011). In particular, females tend to shy away from competitive settings and they are more risk averse than males (see Niederle & Vesterlund, 2007 and Charness & Gneezy, 2012). When faced with a competitive environment, males tend to respond positively, while females do not (Gneezy, Niederle, & Rustichini, 2003; Gneezy & 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 & Pate, 2011; Dargnies, 2012). Moreover, a recent experiment by Kuhn and Villeval (2015) shows that women are more likely than men to enter team-based environments. In light of this, it is interesting to study gender differences in the effect of team RPF. Indeed, 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.

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 & Shaw, 2007; Bandiera, Barankay, & Rasul, 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. [3]

Peer pressure and team spirit is the common explanation for why team incentives work better than standard theory predicts.[4] Alchian and Demsetz (1972) note in their classic book on team production that "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) introduces 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) incorporates 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 not 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, no one has yet studied the effect of relative performance feedback to teams in a laboratory experiment. 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).

[3]A range of studies employing different empirical approaches have identified mixed effects of team incentives. In some field studies, there is an 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, Nickerson, and Owan (2003) and Boning, Ichniowski, and Shaw (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.

[4]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 & Milgrom, 1990; Itoh, 1991, 1992; Macho-Stadler & Pérez-Castrillo, 1993). 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.

A number of experiments have supported this conjecture. Group competition can induce more cooperation (Bornstein and Ben-Yossef, 1994), less free-riding (Bornstein, Erev & Rosen, 1990; Bornstein & Erev, 1994; and Erev, Bornstein, & Galili (1993)) and better coordination (Bornstein, Gneezy, & Nagel, 2002). See in particular Erev et al (1993) who in a field experiment find 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 contributions are compared to another group.[5] This clearly resemble and support 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 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 a conceptual framework. In section 3 we explain our experimental design. In Section 4 we present the results from the experiment, while section 5 concludes.

2. Conceptual Framework

To fix ideas, we present a simple conceptual framework enabling us to present some behavioral predictions. Let there be *n* agents in the economic environment. Agent *i* exerts effort e_i incurring a private cost $c(e_i)$ where $i = 1 \ldots n$, and where the cost function has standard properties $c'(e_i) > 0, c''(e_i) > 0$. He receives a wage $w(e_i, \ldots, e_n)$ and is assumed to have the following utility function:

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

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. Building on Clark and Oswald (1998), 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. This weight can be interpreted as status concerns.

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

In addition, we add a peer pressure function *P*, similar to Kandel and Lazear (1992). Peer pressure is social and/or moral costs, 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. But 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 distinguishes 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. Now, if we introduce peer pressure, 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 now that agents have competitive preferences and care about comparisons. 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. However, there are two reasons why team RPF may have a different effect than individual RPF in our framework. First, agents may 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, with team RPF, peer pressure also works 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 without peer pressure. This might not be fully possible, but it is natural to assume that the lower the peers' effort, the lower is the peer pressure to work hard. 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 when the agent works on behalf of the team (as, say, team leader) and not only when he works along with other team members.

In our framework, heterogeneous responses to RPF can also give insights 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 then have a stronger effect on low performing agents if peer pressure is important. Moreover, differences in ability and/or performances within a team does affect rank utility *v* in our specification. Hence, if one observe higher team RPF response from the top performers within teams, the plausible explanation would be that the weight on status concerns, θ, differs between the agents.

Finally, 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 from 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 *both* have team incentives and team feedback, compared to when only one of the features is in place. The second mechanism is via status concerns. As noted above, agents may put different weight on *v* when it is about team comparisons rather than individual comparisons. 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.

3. Experimental Design

3.1 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.

3.2 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.[6] In any treatment, subjects always learn their individual absolute performance. In any team treatment, subjects always 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 individual (ind) or team (team) incentives, and finally whether the level of feedback is on individuals (ind) or teams (team).

[6]We do not display any aggregate information based on several previous stages.

We introduce treatments gradually. We start by keeping one dimension fixed, and only present treatments that contain RPF. These are displayed in Table 1 and then explained below.

Table 1: Summary of RPF treatments

RELATIVE PERFORMANCE FEEDBACK	Individual RPF	Team RPF
Individual incentive	RPF-ind-ind	RPF-ind-team
Team incentives	RPF-team-ind	RPF-team-team

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. Their performance is ranked (from 1 to 3) and they learn 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. The team RPF consists of performance information about two other teams in the session. 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 the two other team members.[7] Their performance is ranked (from 1 to 3) and they learn how many tasks 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.

RPF-ind-ind and RPF-team-ind are referred to as individual RPF treatments.

In the *RPF-team-team* treatment, subjects receive both team RPF and team incentives, rather than individual RPF and individual incentives. 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 one third of the total number of tasks their team solves.

[7]We choose to provide intra group individual RPF to keep the setup somewhat realistic. 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.

Finally, we introduce our "baseline" conditions, where we do the same variations as in the previous table, only with APF instead of RPF. These are displayed in Table 2 and explained below.

Table 2: Summary of APF treatments

ABSOLUTE PERFORMANCE FEEDBACK	Individual APF	Team APF
Individual incentive	APF-ind-ind	
Team incentives		APF-team-team

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. 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.

We have not collected data for the two cells left empty in Table 2, as the primary use of APF treatments is to establish "baseline" performances. Thus, we have only included APF treatments that are of main interest to compare with RPF treatments. The empty cells are also less realistic. For example, in an *APF-team-ind* treatment, subjects would only receive individual performance feedback, but then it makes no sense to make their earnings depend on other (unknown) team members. Notice also that all treatments actually include APF, and hence RPF is an additional piece of information in the RPF treatments.

3.2.1 Team leader treatments

Our conceptual framework 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. In these "team leader" treatments we have removed (or at least 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 about the performance of the team leader. Incentives are team-based: 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 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, the team leader's performance is ranked (from 1 to 3) and they learn during each break how many tasks two other team leaders have 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. The team RPF consists of performance information about two other team leaders in the session.

In order to highlight the team leader role, and to minimize team spirit effects, we let the passive team members only see the team leader's performance at the end, not during each break.[8] This also allowed for a simpler procedure: In each session, after the working period, the team leaders were told that they have also been a passive member of two other teams. They then learn the performance of their team leaders, and their final earnings are based on their own performances as team leader and two others' performance as team leaders, in addition to the show-up fee.

3.3 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.[9] 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.[10] 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

[8]Admittedly, this was not made explicitly clear to the team leaders, so the team leader might have been under the impression that the team members got feedback about the team leaders performance 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.

[9]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 A-1.

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

of Arts and Education. 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 minutes. 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 527 subjects participated in the experiment. We aimed for 60 subjects in each of the first six treatments and 90 subjects in each of the two team leader treatments. Due to a combination of overbooking and no show, we ended with the following number of participants in each respective treatment: 68 (29 females, 39 males) in APF-ind-ind, 55 (22 females, 33 males) in RPF-ind-ind, 53 (16 females, 37 males) in RPF-ind-team, 56 (23 females, 33 males) in APF-team, 54 (27 females, 27 males) in RPF-team-team, 63 (27 females, 36 males) in RPF-team-ind, 93 (50 females, 43 males) in APF-teamleader and 84 (49 females, 35 males) in RPF-teamleader.[11]

4. Experimental Results

In the section, we present our experimental results. We first present the main treatment effects. Then we study interaction effects, heterogeneous effects, and gender effects in separate sections. Finally, we present results from the two team leader treatments.

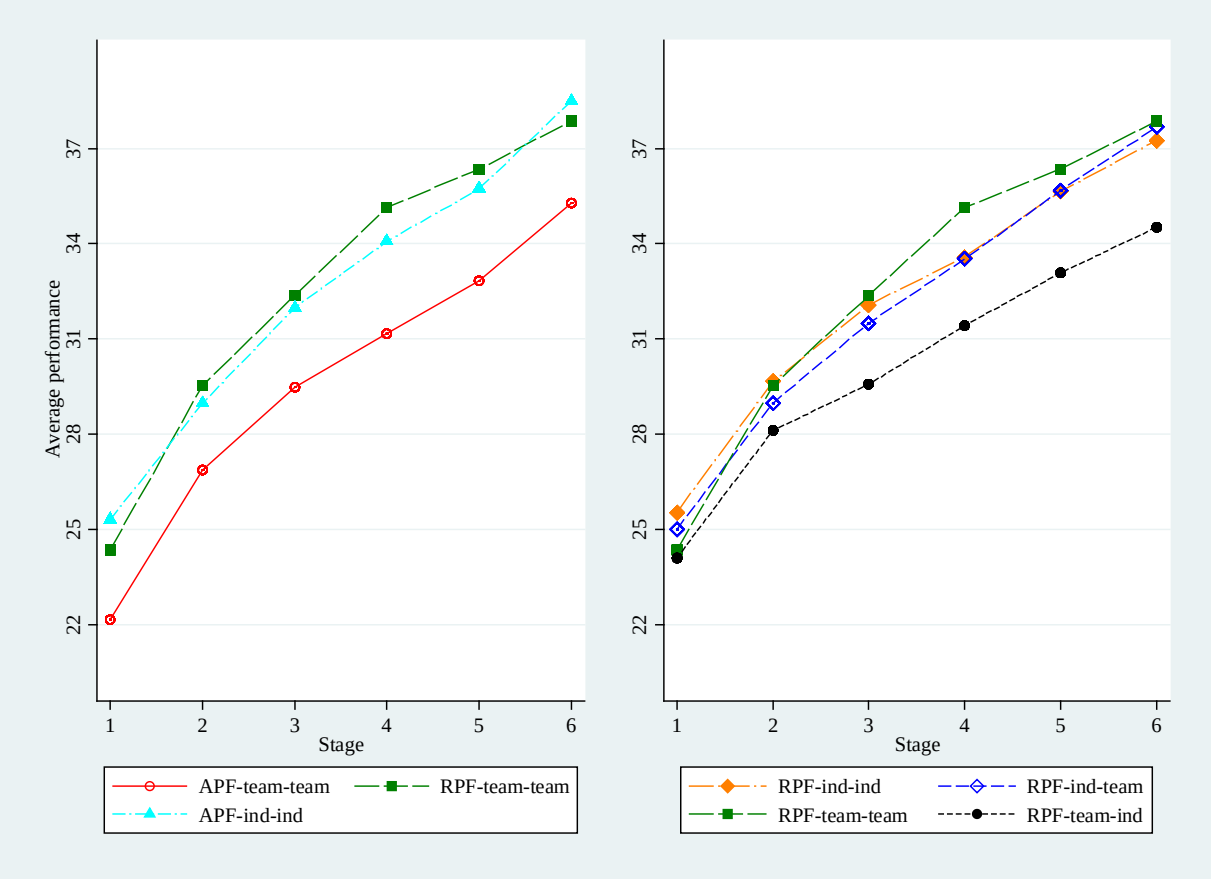
In the regression analysis, we study the 1st and 2nd stage separately. The 1st stage is a "kick-off" stage, as any treatment effect of RPF is driven by the knowledge about future feedback, and not a response to the feedback itself (as found in e.g., Blanes i Vidal & Nossol, 2011). The 2nd stage is the first working stage after any feedback is provided, and the cleanest way to identify any treatment effects of RPF.

[11]Administrative revision found that three subjects participated twice (disregarding explicit information about this being strictly prohibited), two in RPF-ind and one in RPF-ind-team. These subjects are not part of the given number of participants. In addition, these subjects could have affected their peer groups (2+2 subjects in RPF-ind and 8 subjects in RPF-ind-team). We still include these subjects, but results are robust to excluding them.

4.1 Main observations

Figure 1 displays average performance of subjects across stages.

Figure 1: Average performance across stages



We first compare the performance of subjects in RPF-team-team to subjects in APF-team-team. From Figure 1 we see a clear treatment effect. The average performance in RPF-team-team (32.6 tasks solved) is significantly greater (Mann Whitney U-test (MW): p=0.09, Randomization test (RT): p=0.02) than in APF-team-team (29.6), see Table 3. [12,13,14] . The performance is about 10% higher in RPF team-team

[12]We use Mann-Whitney U-test and Randomization test when comparing means throughout this section, unless otherwise specified. When based on the performance across all stages, we use each subject's average performance across all of these. Notice that we do not use a cluster version of MW, however we do compare team averages (see footnote 15). The Randomization tests are based on 200.000 simulations.

[13]The difference (26.9 vs. 29.5) in the 2nd stage is also significant, MW: p=0.04 and RT: p=0.01.

[14]We use the Stata program *permtest2* by Kaiser (2007) 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-

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 higher effort. This is consistent with subjects having relative concerns.[15] Those who receive rank information on team performance increase their effort even if there is no monetary incentives related to rank.

Table 3: Team incentives and RPF

	Average Performance (SD)			Mann-Whitney z-Statistics (p-value)	
	APF-team-team	RPF-team-team	RPF-team-ind		
	(1)	(2)	(3)	(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: Mann-Whitney pairwise test. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Next, observe that performance of subjects receiving individual RPF under team incentives perform at the same level as subjects receiving team APF.[16] Moreover, the direct comparison between the two RPF treatments under team incentives provides significant differences from stage 2 and onward (MW: p=0.09, RT: p=0.03), in that subjects receiving team RPF perform better.[17] 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.

parametric estimation approaches. Moreover, several researchers have recently discussed the use of Randomization test in experimental papers (e.g., Imbens & Rubin, 2015; Young, 2017).

[15]A different approach is to compare team averages rather than subject averages. In such analysis, the difference between APF-team-team and RPF-team-team over all periods is even more significant with p=0.026 (based on 38 observations).

[16]In this comparison there is only one condition that changes. As subjects in the RPF-team-ind learn everything that subjects in the APF-team-team learn, the only change is the additional individual RPF.

[17]Including stage 1 leads to an insignificant difference (MW: p=0.13, RT: p=0.05), but considering the development in performance seen in Figure 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.

Consider then the effect of team incentives. Comparing the two APF treatments,[18] we see that the average performance in APF-ind-ind (32.4) is significantly higher (MW: p=0.01, RT: p=0.01)[19] than in APF-team-team (29.6), see Table 4. This is consistent with the free-rider problem discussed in Section 2, as subjects working under individual incentives solve, on average, almost 10% more tasks than those working under team incentives. It is also consistent with previous empirical findings of free-riding activity in teams (see e.g., Corgnet et al., 2015).

Table 4: Free-rider problem

	Average Performance (SD)		Mann-Whitney z-Statistics
	APF-ind-ind	APF-team-team	(p-value)
	(1)	(2)	(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: Mann-Whitney pairwise test. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

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 A-2. Hence, moving from APF-ind-ind to APF-team-team (step 1) revealed a free-rider problem. Moving from APF-team-team to RPF-team-team (step 2) revealed a positive team feedback effect. The net result of these two steps cancel each other out, so that the addition of team RPF (step 2) seems to offset the free-rider problem with team incentives (step 1).

Next, we observe no average performance difference between APF and any RPF under *individual incentives.* The average performance in RPF-ind-ind (32.3) is statistically the same (MW: p=0.42, RT: p=0.91) as in APF-ind-ind (32.4), see Table 5. Moreover, the average performance in RPF-ind-team (32.1) is statistically the same (MW: p=0.58, RT: p=0.79) as in APF-ind-ind (32.4). Hence, the positive effect of *team* RPF applies only under team incentives, not under individual incentives. Moreover, subjects under

[18]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.

[19]If we only study the difference in stage 2 (29.0 vs. 26.9), the difference is also significant with MW: p=0.02 and RT: p=0.02. Using team averages, as described in footnote 15, the difference over all periods is also significant with p=0.090.

individual incentives perform equally well, independently of the type of feedback provided. This is somewhat surprising, given both the theoretical expectation of a motivational effect from RPF and the empirical finding that RPF indeed matters in team settings. However, also other have found no effect of individual RPF (e.g. Eriksson et al, 2009).

Table 5: Individual incentives and RPF

	Average Performance (SD)			Mann-Whitney z-Statistics (p-value)	
	APF-ind-ind	RPF-ind-ind	RPF-ind-team		
	(1)	(2)	(3)	(1) vs (2)	(1) vs (3)
Stage 1	25.31 (4.90)	25.53 (5.91)	25.00 (4.93)	0.29 (0.771)	0.52 (0.606)
Stage 2	28.97 (5.32)	29.65 (6.11)	29.04 (5.27)	-0.17 (0.862)	-0.06 (0.954)
All stages	32.43 (6.06)	32.29 (6.95)	32.13 (5.84)	0.80 (0.423)	0.56 (0.576)
N	68	55	53	123	121

Notes: Mann-Whitney pairwise test. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Thus far, we base our observations on comparing mean performances, not controlling for any other potentially important characteristics.[20] Reported in Table 6 are OLS and Random Effects GLS estimations, controlling for other factors such as age and gender.[21,22] 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.[23]

[20]In the appendix, Table A-1, we check for randomization across treatments. Some minor differences exist, so controlling for such differences may prove important to the robustness of our findings.
[21]In the regressions, we use robust standard errors clustered on sessions. However, as the number of clusters may be too low, it could downward bias our 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.
[22]Alternatively, we could increase 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.
[23]The remaining stages are in the appendix, Table A-3.

Table 6: Main results: Treatment effects on productivity

Stage(s):	1st stage	2nd stage	All stages
	(1)	(2)	(3)
APF-team-team	Ref.	Ref.	Ref.
APF-ind-ind	3.149***	2.202***	2.529***
	(0.8345)	(0.2976)	(0.6027)
RPF-ind-ind	3.758**	2.887***	3.563***
	(1.3377)	(0.5186)	(0.8071)
RPF-ind-team	3.471***	2.733***	3.559***
	(1.1521)	(0.8012)	(0.7455)
RPF-team-team	2.618	2.720**	3.578**
	(1.8772)	(1.0981)	(1.2679)
RPF-team-ind	2.437*	1.510**	1.426**
	(1.3011)	(0.5784)	(0.6648)
$Stage_t$			2.366***
			(0.0757)
Constant	31.428***	35.059***	32.604***
	(2.9264)	(2.5666)	(2.6599)
Adjusted R^2	0.094	0.059	
Observations	350	350	2100

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.

$^{*}\ p < 0.10,\ ^{**}\ p < 0.05,\ ^{***}\ p < 0.01.$

First, in column (3), we observe the significant effect of RPF-team-team relative to the baseline. The effect is consistent throughout the working stages. We can then establish our first main result:

Result 1: *Subjects working under team incentives perform on average significantly better when they receive team RPF, i.e. feedback on how their team is doing compared to other teams*

This result is in accordance with theory presented in Section 2 and complement Delfgaauw et al. (2013) who find a similar result in a field experiment.

Then consider individual RPF under team incentives. The coefficient estimate is positive and significant when we include controls to our estimation, and the performance of subjects in this treatment is slightly higher than the baseline. The effect weakens in the later stages, as can be seen in Table A-3. Compare then the performance in RPF-team-team to the RPF-team-ind. Across all stages the RPF-team-team subjects outperform the RPF-team-ind subjects by about 2 tasks, but this difference is close to significance at the 10% level (p=0.101). However, if we only look at stages 2-6, allowing subjects to respond to the feedback, the difference between them is significant (p=0.036, see also Table A-4 in the appendix).

The regressions also establish the presence of free-riding in teams, and moreover that subjects under individual incentives are less responsive to feedback regimes. However, we do find that subjects in the RPF-ind-team perform significantly better than in the APF-ind-ind (p=0.07), revealing a slightly positive effect of team RPF on performance also under individual incentives.

From Table A-4, 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

The RPF treatments fit into a 2 x 2 design, varying between individual incentives or team incentives and individual RPF or team RPF (see Table 1).[24] 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

[24]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).

in Table 6. 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 7 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 are complements, i.e. providing team RPF positively strengthens the influence of team incentives, and vice versa.

Table 7: Changing incentives and feedback

Stage(s):	All stages	Stages 1-3	Stages 4-6
	(1)	(2)	(3)
Individual incentives and individual RPF	Ref.	Ref.	Ref.
Team incentives	-2.694***	-2.563***	-2.826***
	(0.6718)	(0.6420)	(0.7486)
Team RPF	-0.456	-1.022	0.109
	(0.5998)	(0.5914)	(0.6856)
Team incentives x Team RPF	3.533***	3.484***	3.583***
	(0.9364)	(0.9474)	(0.9909)
Stage $_t$	2.301***	3.300***	1.698***
	(0.0992)	(0.1719)	(0.1337)
Constant	32.958***	28.822***	38.110***
	(2.9674)	(2.4412)	(3.8004)
Observations	1350	675	675

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. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Result 2: *Team incentives and team RPF are complements.*

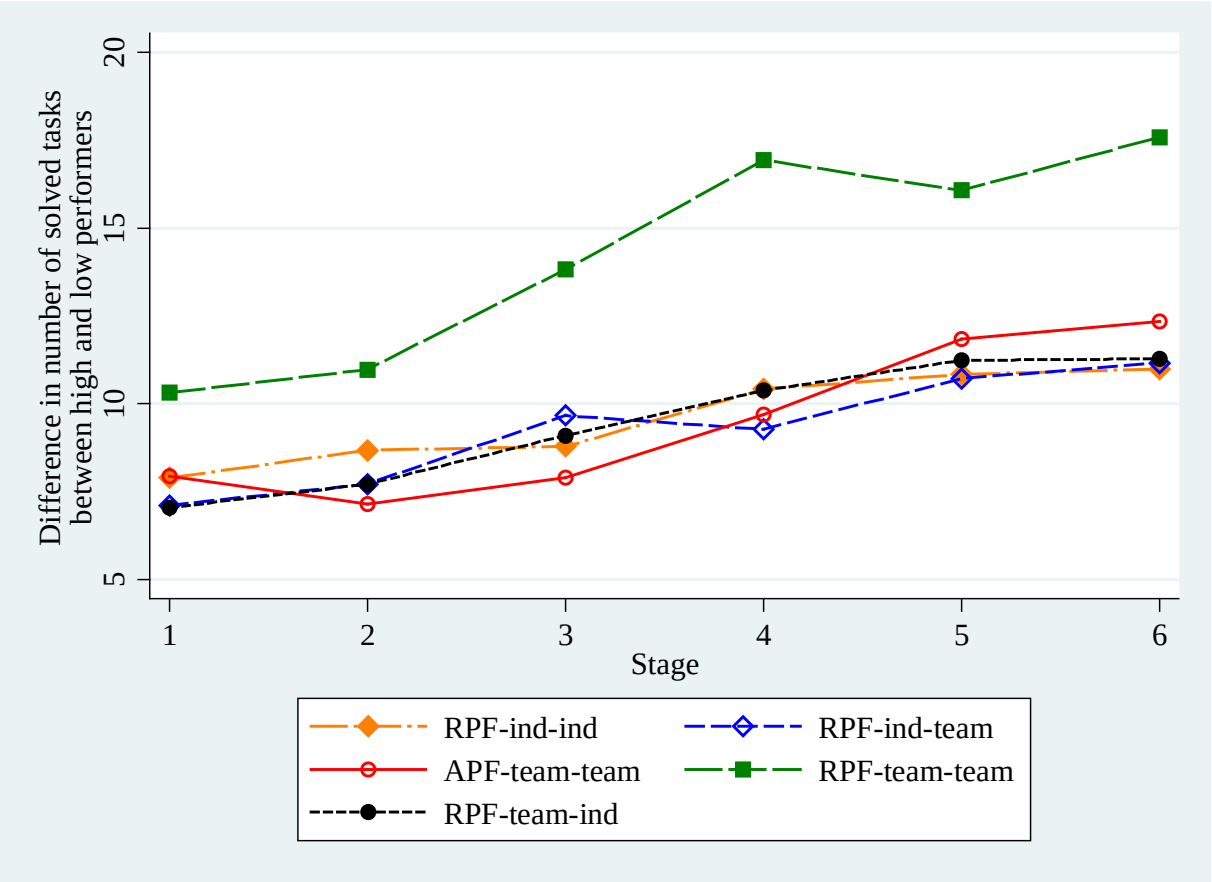
From the theory presented in Section 2, this result can be explained by peer pressure or status concerns (or both). If it is peer pressure, then the marginal positive effect of effort from agent i on agent j's utility is reinforced when agents have both team incentives and team RPF. If it is status concerns, then agents put higher weight on rank under team RPF then under individual RPF, and in particular so when agents also are exposed to team incentives. Results from the coming sections will help us disentangle these two explanations.

4.3 Heterogeneous effects

We have seen that RPF affects average performances. In this section, we investigate heterogeneous effects, i.e. to what extent the treatments affect the performance distributions.

We categorize subjects within a team as either best or worst, based on their average performance over all stages. Hence, a subject categorized as best keeps this categorization in all rounds, even though someone else in the team may have done better in a single stage. We compare the difference between the performance of the best and the worst subject within a team, and compare this difference across treatments. Figure 2 shows a substantially larger gap between the best and the worst performers within a team in the RPF-team-team, compared to any other treatment. 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. Who drives the difference that we see in Figure 2? Figure A-1 shows that high performers in the RPF-team-team perform better than high performers in other treatments.

Figure 2: Difference between high and low performers across treatments



Quantile regressions reported in Table A-5 support these findings. It is in particular the highest performers in the RPF team feedback treatments that perform better than the highest performers in the other treatments.

Result 3: *High performing subjects perform better in treatments with team RPF than high performing subjects in any other treatment.*

In Table 8, we present regressions including a dummy variable (BiT) for the subject who is the best performer within the team. This variable is also interacted with each treatment. Hence, the sum of the coefficients BiT and the treatment interacted with BiT, is the additional tasks the best performer solved relative to the other two subjects within the team. Therefore, to compare best performers within a team across treatments, say between best performers in RPF-ind-ind and APF-team-team, one has to take the

difference between them. That is, for the concrete example, one has to sum the coefficients for RPF-ind-ind and RPF-ind-ind x BiT in order to find the corresponding estimated difference.[25,26]

Consistent with Figure A-1, best performers in RPF-team-team perform significantly better than best performers in the baseline. Moreover, the best performers in RPF-team-team also perform significantly better than the best performers in both RPF-ind-team (p=0.03) and RPF-team-ind (p=0.01).[27] Notably, in Table A-5, high performers seemed to perform better in both team RPF treatments. Table 8 suggests that for RPF-team-team this is driven by the best performer within each team, whereas in RPF-ind-team it must be that there are more often multiple high performers in the same team (as there is no interaction effect between BiT and RPF-ind-team).

This implies that team incentives motivate significantly higher top performance than individual incentives, when subjects are exposed to team RPF. Notice, however, that the second and third performers within the team perform significantly worse under team incentives than under individual incentives. This can be seen directly from the coefficients to RPF-ind-ind and RPF-ind-team when compared against the baseline, but the difference is also significant for those in RPF-team-ind relative to the individual incentive treatments (column (3), both p<0.01).

Moreover, differences in ability and/or performances within a team do affect rank utility *v* in our theoretical framework. 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 the agents. Moreover, it suggests that peer pressure is not as influential in this setting, as this would imply a stronger response from low performing subjects.

Result 3 suggests that subjects' weight on status concerns differ between the agents. 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).

[25]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), i.e. (2.05+4.30) – (1.45-0.10) = 5.
[26]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, i.e. the statistical power is reduced.
[27]Also in point estimates against the best performers in RPF-ind-ind (p=0.14).

Table 8: Best performers across treatments

Stages:	1st stage	2nd stage	All stages
	(1)	(2)	(3)
APF-team-team	Ref.	Ref.	Ref.
RPF-ind-ind	3.501**	2.222***	3.540***
	(1.5584)	(0.6901)	(0.7625)
RPF-ind-team	3.466**	2.525**	3.647***
	(1.5296)	(1.0740)	(0.8219)
RPF-team-team	1.875	1.458	2.052
	(1.9212)	(0.8831)	(1.2320)
RPF-team-ind	2.610	1.229	1.453*
	(1.5699)	(0.7455)	(0.6980)
BiT (Best in Team)	5.481***	5.047***	6.864***
	(1.2606)	(1.2029)	(0.8641)
RPF-ind-ind x BiT	0.416	1.356	0.007
	(1.7578)	(1.7810)	(1.5093)
RPF-ind-team x BiT	-0.322	0.045	-0.447
	(1.5786)	(1.2655)	(1.0078)
RPF-team-team x BiT	2.192	3.577*	4.297***
	(1.7282)	(1.6955)	(1.2533)
RPF-team-ind x BiT	-0.849	0.357	-0.104
	(1.4659)	(1.4605)	(1.0329)
$Stage_t$			2.328***
			(0.0835)
Constant	25.553***	29.417***	25.651***
	(2.6936)	(1.9901)	(2.2125)
Adjusted R^2	0.295	0.321	
Observations	282	282	1692

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.

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

As a remark, it should be noted that we do not have independent ability measures in our study, i.e. a measure of the ability to solve coding tasks that are independent of treatments. However, we can use University grades as a proxy for more general ability, and use it to study whether people with different ability respond differently to relative performance feedback. We categorize an average grade of C or below as low ability, whereas B or above is categorized as high ability. Grades turns out to correlate positively with performance in our experiment, but not significantly. Interestingly, we find that high ability subjects respond positively to team RPF, while low ability subjects do not. However, the interaction effect between ability and response to team RPF is not significant. More results on differential response can be found in Table A-6 column 1 in the appendix.

4.4 Gender analysis

In this section, we study gender effects. In Table 9, we add gender indicators interacted 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. Observe that the performance of males is very much in line with the overall results. Under individual incentives, males in RPF-ind-ind outperform males in both APF-ind-ind ($p<0.05$) and RPF-ind-team ($p<0.10$), suggesting a strong motivational effect of individual feedback. Under team incentives, males in RPF-team-team ($p<0.01$) and RPF-team-ind ($p<0.10$) outperform males in APF-team-team, and males in RPF-team-team do better than males in RPF-team-ind ($p<0.10$). There is no significant difference for females under individual incentives, but females in RPF-team-team outperform females in APF-team-team and RPF-team-ind. See also Figure A-2 and Figure A-3 in the appendix.

Table 9: Gender analysis

	All stages	Stages 2-6
	(1)	(2)
APF-team-team	Ref.	Ref.
APF-ind-ind	2.193**	-0.257
	(0.9298)	(0.6963)
RPF-ind-ind	4.222***	0.058
	(0.4329)	(0.9101)
RPF-ind-team	2.700***	0.109
	(0.7755)	(0.6812)
RPF-team-team	3.658***	0.614
	(1.2952)	(0.7037)
RPF-team-ind	1.000*	-1.169
	(0.5405)	(0.7247)
Female	-1.181	0.271
	(1.4218)	(0.5235)
APF-ind-ind x Female	0.611	-0.724
	(2.1984)	(0.9671)
RPF-ind-ind x Female	-2.040	-1.815**
	(1.4366)	(0.8441)
RPF-ind-team x Female	2.798	-0.436
	(3.3482)	(1.4932)
RPF-team-team x Female	0.179	1.335
	(1.8631)	(1.2666)
RPF-team-ind x Female	0.387	-1.211
	(1.4968)	(0.8536)
$Stage_t$	2.009***	2.009***
	(0.0766)	(0.0766)
Observations	1750	1750

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. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Turn then to gender differences. Males strongly outperform females in RPF-ind-ind ($p<0.001$) and are close to outperforming them in RPF-team-ind ($p=0.12$). There are no other significant gender differences between males and females, suggesting that females only dislike relative feedback when it involves individual relative feedback. Actually, females do better than males in RPF-ind-team, although this difference is not significant.

Next, we use the first stage as control, to see how the treatments develop differently after the first stage. Although possibly endogenous, we see that females significantly worsen their already low performance in the RPF-ind-ind ($p<0.01$) and RPF-team-ind ($p<0.01$) relative to females in APF-team-team. The development of females in RPF-ind-ind is significantly negative relative to the development of males in RPF-ind-ind ($p<0.05$), further strengthening the gender difference after the first stage in this treatment. Moreover, females significantly improve their already positive performance in RPF-team-team ($p<0.05$).

In sum, these observations support previous findings (as in Azmat & Iriberri, 2016):

Result 4: *Males respond positively to individual RPF, while females respond negatively. Both genders respond positively to team RPF. Individual RPF makes females produce less.*

In Table A-6, column (3) in the appendix, we present interactions between treatments, ability and gender. It shows that almost the entire gender difference in RPF-ind-ind is due to low ability females performing statistically worse than low ability males, [28] whereas there are no gender differences for those with high ability.[29] Low ability females in RPF-ind-ind actually perform significantly ($p<0.01$) worse than low ability females in APF-ind-ind.

Consider now the interaction effects between feedback and incentives. In Table A-7 we employ the same analysis as in Section 4.3, but on each gender separately. First, observe that males respond more negatively to team incentives alone than females. Second, males respond negatively to team RPF alone, whereas females respond positively. Hence, while males are triggered by individual RPF, females are triggered by team RPF. Finally, we observe that the positive interaction effect demonstrated in Table 7 is gender specific. For males there is a strong complementarity between team incentives and team RPF, although the net effect of shifting both factors is insignificant. Females, on the other hand, only need team RPF to improve performance, and do not gain additional productivity when interacting the two variables.

[28]To find the difference between low ability males and low ability females in this treatment: Female + RPF-ind-ind x Female + Low x Female + RPF-ind-ind x Low x Female, which is -6.06 and $p=0.030$.
[29]As we separate on both gender and ability, the number of observations in each cell is lower; this calls for a cautious interpretation of the results.

Their net differential performance of changing to both team incentives and team RPF (the sum of all coefficients) is positive (p=0.047).

Result 5: *Females respond positively to team RPF, independently of incentives. Males respond negatively to both team incentives and team RPF alone, but a strong positive complementary effect between the two offsets the negative effects.*

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. We have also shown that team incentives and team RPF are complements. It remains to identify the mechanism behind these results. In the theory section, 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 is not a main explanation of our results. Our team leader treatments are meant to further explore this. The approach is to reduce peer pressure by letting people *work on behalf of teams as team leaders,* where the others in the team do not work.

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 10, we run regressions and find the effect in the 2nd stage to be significant at the 5% level. Figure A-4 graphs the development in performance across stages for both treatments.

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 exactly 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 *framin*g may in itself be motivating.

Table 10. Team leader results: Effects on productivity

Stage(s):	1st stage	2nd stage	All stages
	(1)	(2)	(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.

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

Result 6: *Subjects who work as team leaders and receive team RPF outperform subjects who work as team leaders and only receive team APF. The effect is statistically significant in the first working periods, but not in the last periods.*

This result, together with the top performer result (Result 3), 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 RPF effect is reduced during the team leader treatments. 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 percent. 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. In fact, team incentives trigger significantly higher top performance than individual incentives, when subjects are exposed to 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.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly journal of Economics, 115*(3), 715-753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. *The Journal of Economic Perspectives, 19*(1), 9-32.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review, 62*(5), 777-795.

Auriol, E., & Renault, R. (2008). Status and incentives. *The RAND Journal of Economics, 39*(1), 305-326.

Azmat, G., & Iriberri, N. (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics & Management Strategy, 25*(1), 77-110.

Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015). Letting Down the Team? Social Effects of Team Incentives. *Journal of the European Economic Association, 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. *Academy of Management Journal, 53*(4), 827-845.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association, 11*(5), 1079-1114.

Barankay, I. (2012). *Rank incentives: Evidence from a randomized workplace experiment*. Working paper. The Wharton School.

Bellemare, C., Lepage, P., & Shearer, B. (2010). Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace. *Labour Economics, 17*(1), 276-283.

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

Birkinshaw, J. (2001). Why is knowledge management so difficult? *Business Strategy Review, 12*(1), 11-18.

Blanes i Vidal, J., & Nossol, M. (2011). Tournaments without prizes: Evidence from personnel records. *Management Science, 57*(10), 1721-1736.

Boning, B., Ichniowski, C., & Shaw, K. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics, 25*(4), 613-650.

Bornstein, G. & Erev, I. (1994) The enhancing effect of intergroup competition on group performance *International Journal of Conflict Managemen*t, 5, 271-281

Bornstein, G,; Erev I., & Rosen, O (1990) Intergroup competition as a structural solution to social dilemmas. *Social Behaviou*r, 5, 247-260

Bornstein, G., & Ben-Yossef, M. (1994). Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology, 30*(1), 52-67.

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior, 41*(1), 1-25.

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

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. *Journal of Economic Behavior & Organization, 83*(1), 50-58.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization, 45*(3), 301-328.

Charness, G., Masclet, D., & Villeval, M. C. (2014). The dark side of competition for status. *Management Science, 60*(1), 38-55.

Che, Y.-K., & Seung-Weon, Y. (2001). Optimal incentives for teams. *The American Economic Review, 91*(3), 525-541.

Chen, H., & Lim, N. (2013). Should Managers Use Team-Based Contests? *Management Science, 59*(12), 2823-2836. doi:10.1287/mnsc.2013.1743

Clark, A. E., & Oswald, A. J. (1998). Comparison-concave utility and following behaviour in social and economic settings. *Journal of Public Economics 70, 133–155*.

Corgnet, B., Hernán-González, R., & Rassenti, S. (2015). Peer Pressure and Moral Hazard in Teams: Experimental Evidence. *Review of Behavioral Economics, 2*(4), 379-403. doi:10.1561/105.00000040

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature, 47*(2), 448.

Dargnies, M.-P. (2012). Men too sometimes shy away from competition: The case of team competition. *Management Science, 58*(11), 1982-2000.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament Incentives in the Field: Gender Differences in the Workplace. *Journal of Labor Economics, 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. *Journal of Experimental Social Psychology, 29*(6), 463-478.

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics, 16*(6), 679-688.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171-178.

Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status*: Oxford University Press.

Gjedrem, W. G. (2015). *Relative performance feedback: Effective or dismaying?* Working paper. UiS Business School. University of Stavanger, Norway.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics, 118*(3), 1049-1074.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *The American Economic Review, 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. *American Economic Journal: Applied Economics, 1*(4), 34-68.

Hamilton, Barton H., Nickerson, Jack A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy, 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. *The Accounting Review, 83*(4), 893-913.

Healy, A., & Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal, 121*(555), 1192-1204.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324-340.

Holmström, B., & Milgrom, P. (1990). Regulating trade among agents. *Journal of Institutional and Theoretical Economics*, 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. *Journal of Law, Economics, & Organization, 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 Journal, 7*(3), 402-412.

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy, 100*(4), 801-817.

Knez, M., & Simester, D. (2001). Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics, 19*(4), 743-772.

Kuhn, P., & Villeval, M. C. (2015). Are Women More Attracted to Co-operation Than Men? *The Economic Journal, 125*(582), 115-140.

Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science, 58*(1), 94-113.

Kvaløy, O., & Olsen, T. E. (2006). Team incentives in relational employment contracts. *Journal of Labor Economics, 24*(1), 139-169.

Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy, 89*(5), 841-864.

Lazear, E. P., & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives, 21*(4), 91-114.

Macho-Stadler, I., & Pérez-Castrillo, J. D. (1993). Moral hazard with several agents: The gains from cooperation. *International Journal of Industrial Organization, 11*(1), 73-100.

Marino, A. M., & Zábojnik, J. (2004). Internal Competition for Corporate Resources and Incentives in Teams. *The RAND Journal of Economics, 35*(4), 710-727.

Nalbantian, H. R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *The American Economic Review, 87*(3), 314-341.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics, 122*(3), 1067-1101.

Sausgruber, R. (2009). A note on peer effects between teams. *Experimental Economics, 12*(2), 193-201.

Tan, J. H., & Bolle, F. (2007). Team competition and the public goods game. *Economics Letters, 96*(1), 133-139.

van Dijk, F., Sonnemans, J., & van Winden, F. (2001). Incentive systems in a real effort experiment. *European Economic Review, 45*(2), 187-214.

Vandegrift, D., & Yavas, A. (2011). An experimental test of behavior under team production. *Managerial and Decision Economics, 32*(1), 35-51.

Young, A. (2017). *Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*. Working paper. London School of Economics.

Appendix

Table A-1: Summary statistics of control variables

	APF-ind-ind	RPF-ind-ind	RPF-ind-team	APF-team-team	RPF-team-team	RPF-team-ind	Pearson^2 /Kruskal Wallis	APF-team-leader	RPF-team-leader	Pearson^2/ Kruskal Wallis
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Economics-students	0.132 (0.341)	0.036 (0.189)	0.038 (0.192)	0.123 (0.331)	0.204 (0.407)	0.143 (0.353)	0.044[30]	0.226 (0.420)	0.083 (0.278)	0.009
Norwegian-Nationality	0.706 (0.459)	0.473 (0.504)	0.434 (0.500)	0.579 (0.498)	0.519 (0.504)	0.413 (0.496)	0.010[31]	0.559 (0.499)	0.560 (0.499)	0.996
Age	24.29 (4.316)	26.05 (3.955)	26.08 (4.751)	24.25 (3.291)	25.57 (4.364)	25.35 (4.656)	0.015	26.04 (7.228)	25.37 (4.935)	0.879
Female	0.426 (0.498)	0.400 (0.494)	0.302 (0.463)	0.404 (0.495)	0.500 (0.505)	0.476 (0.503)	0.365	0.538 (0.501)	0.583 (0.496)	0.541
Average-grade	2.559 (0.720)	2.055 (0.780)	2.340 (0.678)	2.526 (0.782)	2.370 (0.623)	2.508 (0.592)	0.003[32]	2.559 (0.787)	2.500 (0.768)	0.368
Observations	68	55	53	57	54	63	350	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).

[30]Excluding RPF-ind leads these differences to be insignificant (p=0.150)
[31]Excluding APF-ind leads these differences to be insignificant (p=0.385)
[32]Excluding RPF-ind leads these differences to be insignificant (p=0.352)

Table A-2: Team RPF eliminates free-riding

	Average Performance (SD)			Mann-Whitney z-Statistics (p-value)	
	APF-ind-ind (1)	RPF-ind-ind (2)	RPF-team-team (3)	(1) vs (3)	(2) vs (3)
Stage 1	25.31 (4.90)	25.53 (5.91)	24.35 (6.81)	1.20 (0.230)	0.87 (0.382)
Stage 2	28.97 (5.32)	29.65 (6.11)	29.52 (6.23)	0.01 (0.994)	0.20 (0.841)
All stages	32.43 (6.06)	32.29 (6.95)	32.60 (7.63)	0.46 (0.648)	-0.06 (0.954)
N	68	55	54	122	109

Notes: Mann-Whitney pairwise test. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A-3: Treatment effects across stages

Stages:	3rd stage	4th stage	5th stage	6th stage
	(1)	(2)	(3)	(4)
APF-team-team	Ref.	Ref.	Ref.	Ref.
APF-ind-ind	2.557***	3.070***	2.980***	3.378***
	(0.5777)	(0.6835)	(0.6538)	(0.9001)
RPF-ind-ind	2.726***	2.276***	2.783***	1.778**
	(0.6547)	(0.5267)	(0.7393)	(0.6832)
RPF-ind-team	2.853***	3.539***	3.977***	3.430**
	(0.6817)	(1.1322)	(1.1992)	(1.3356)
RPF-team-team	3.022**	3.882***	3.625***	2.563
	(1.2275)	(1.1901)	(1.1188)	(1.5192)
RPF-team-ind	0.473	0.629	0.780	-0.373
	(0.8411)	(0.6798)	(0.9434)	(1.0629)
Constant	40.523***	45.213***	46.721***	49.436***
	(2.6384)	(2.7256)	(3.1974)	(3.6538)
Adjusted R^2	0.095	0.123	0.109	0.076
Observations	350	350	350	350

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 A-4: Persistence of treatment effects

Stages:	Stages 1-3	Stages 4-6	Stages 2-6
	(1)	(2)	(3)
APF-team-team	Ref.	Ref.	Ref.
APF-ind-ind	2.418***	2.641***	2.406***
	(0.6598)	(0.5627)	(0.5582)
RPF-ind-ind	3.668***	3.457***	3.398***
	(0.9188)	(0.7410)	(0.7196)
RPF-ind-team	3.176***	3.941***	3.509***
	(0.7987)	(0.7189)	(0.6837)
RPF-team-team	3.078**	4.078***	3.771***
	(1.3408)	(1.2122)	(1.1804)
RPF-team-ind	1.804**	1.047	1.099*
	(0.7920)	(0.6119)	(0.5851)
$Stage_t$	3.367***	1.856***	2.009***
	(0.1323)	(0.1062)	(0.0777)
Observations	1050	1050	1750

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 A-5: Marginal treatment effects across quantiles

Quantile:	10%	25%	50%	75%	90%
	(1)	(2)	(3)	(4)	(5)
APF-team-team	Ref.	Ref.	Ref.	Ref.	Ref.
APF-ind-ind	1.396	2.410	2.369	2.711*	1.697
	(1.7573)	(1.4706)	(1.7082)	(1.3899)	(1.9684)
RPF-ind-ind	1.979	2.462	2.482	3.023	4.652*
	(2.1938)	(1.8981)	(1.5015)	(2.0786)	(2.4430)
RPF-ind-team	0.854	2.697	3.321*	4.377**	5.955**
	(1.7065)	(1.6392)	(1.7520)	(1.9019)	(2.2914)
RPF-team-team	-0.417	0.492	1.673	5.545**	7.545***
	(1.7406)	(1.6561)	(1.9975)	(1.9592)	(2.5874)
RPF-team-ind	0.396	0.977	1.470	1.051	2.106
	(2.3174)	(1.8932)	(1.6135)	(1.8456)	(2.2698)
Constant	32.896***	37.541***	36.911***	44.503***	49.333***
	(3.8628)	(4.4789)	(3.7987)	(4.3926)	(3.9050)
Observations	350	350	350	350	350

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. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Figure A-1: High and low performers across treatments

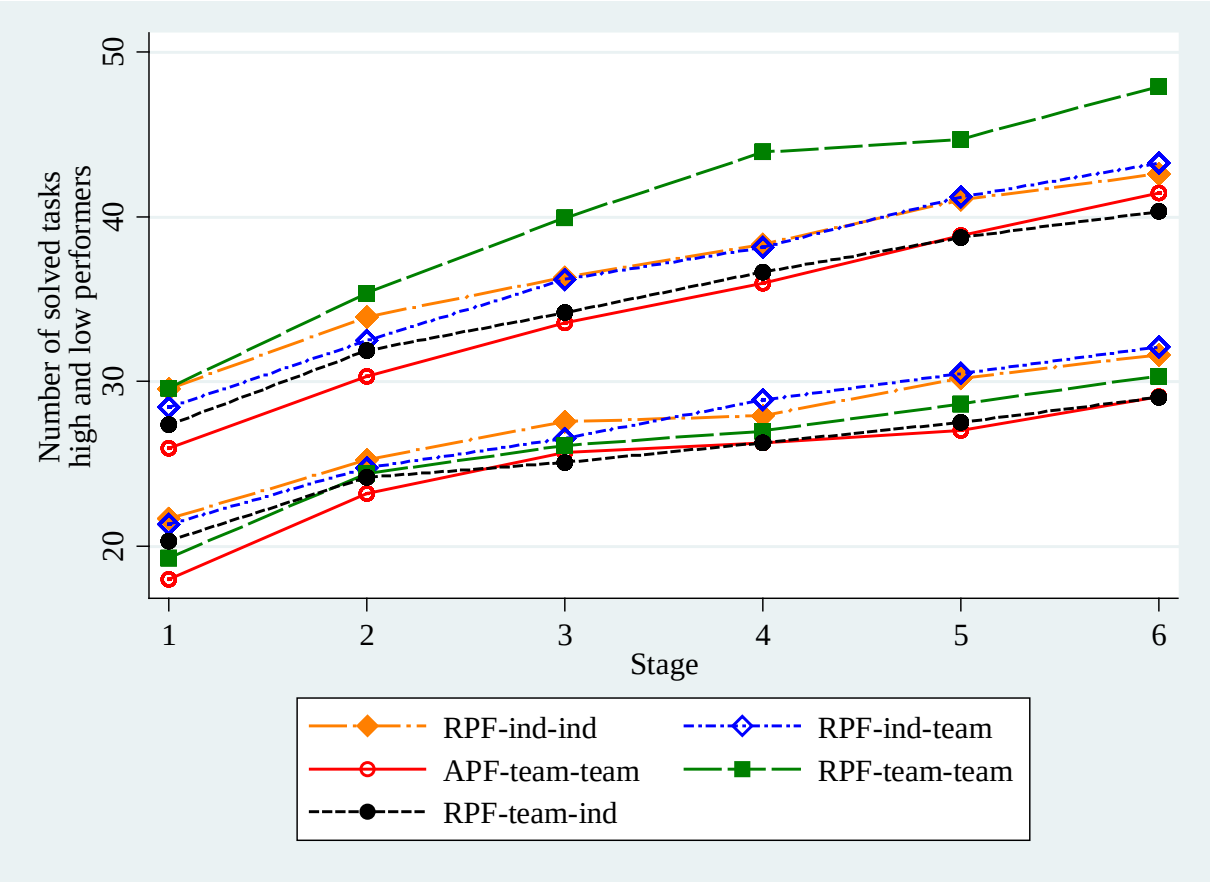


Table A-6: Gender & Ability

	Ability	Gender	Gender & ability
	(1)	(2)	(3)
APF-team-team	Ref.	Ref.	
APF-ind-ind	2.852***	2.193**	1.761
	(0.7184)	(0.9298)	(1.7741)
RPF-ind-ind	3.438***	4.222***	3.181***
	(0.9716)	(0.4329)	(0.6359)
RPF-ind-team	1.501	2.700***	0.789
	(1.1580)	(0.7755)	(1.4278)
RPF-team-team	4.007***	3.658***	5.468***
	(1.3116)	(1.2952)	(1.0960)
RPF-team-ind	-0.878	1.000*	-1.745
	(1.1134)	(0.5405)	(1.3074)
Low	-2.649***		-3.126***
	(0.5486)		(0.3837)
APF-ind-ind x Low	-0.566		0.847
	(0.6271)		(1.5645)
RPF-ind-ind x Low	-2.720		1.166
	(2.5811)		(3.7867)
RPF-ind-team x Low	4.422**		4.805**
	(2.0648)		(2.0324)
RPF-team-team x Low	-0.107		-3.946**
	(3.3694)		(1.6957)
RPF-team-ind x Low	3.369*		4.852**
	(1.7862)		(2.0420)
Female	-0.715	-1.181	-1.756
	(0.6425)	(1.4218)	(1.7720)
APF-ind-ind x Female		0.611	3.011
		(2.1984)	(3.4565)
RPF-ind-ind x Female		-2.040	0.711
		(1.4366)	(2.4783)
RPF-ind-team x Female		2.798	3.383
		(3.3482)	(3.6682)
RPF-team-team x Female		0.179	-2.912
		(1.8631)	(2.2107)
RPF-team-ind x Female		0.387	1.990
		(1.4968)	(2.0011)
Low x Female			1.140
			(1.5458)

APF-ind-ind x Low x Female			-3.701
			(3.3634)
RPF-ind-ind x Low x Female			-6.146
			(3.8046)
RPF-ind-team x Low x Female			-2.611
			(1.9168)
RPF-team-team x Low x Female			7.400**
			(3.5541)
RPF-team-ind x Low x Female			-3.294
			(2.0750)
Stage $_t$	2.009***	2.009***	2.009***
	(0.0766)	(0.0766)	(0.0768)
Constant	33.755***	35.279***	34.310***
	(2.3354)	(2.2883)	(2.2285)
N	1750	1750	1750

Notes: Random Effects GLS coefficients reported, with robust standard errors in parentheses, corrected for clustering across sessions. The dependent variable in all columns is number of solved tasks in all stages. The reference group in column (1) is high ability subjects in APF-team-team, in column (2) it is males in APF-team-team and in column (3) it is high ability males in APF-team-team. In column (3), low ability males is the sum of treatment variable, low and treatment variable x low. High ability females can be found by summing the treatment variable, female and treatment variable x female. Finally low ability females is the sum treatment variable, low, treatment variable x low, female, treatment variable x female, low x female, and treatment variable x female x low. Column (1) focuses on ability, column (2) focuses on gender, and column (3) focuses on their interactions. All columns have the following control variables included: Time on the day of the session (FE in panel), age, a dummy for economics students and a dummy for Norwegian nationality. Column 2 also includes a control for average University grades.

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

Table A-7: Changing incentives and feedback – gender analysis

Panel:	Males	Females
	(1)	(2)
Individual incentives and individual RPF	Ref.	Ref.
Team incentives	-3.600***	-1.326*
	(1.1046)	(0.7177)
Team RPF	-2.652***	3.956**
	(0.8195)	(1.2812)
Team incentives X Team RPF	4.312***	0.124
	(1.2671)	(1.5096)
Stage $_t$	2.355***	2.227***
	(0.1109)	(0.1869)
Constant	37.294***	30.263***
	(4.7561)	(3.1577)
Observations	780	570

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. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Figure A-2: Gender - individual incentive treatments

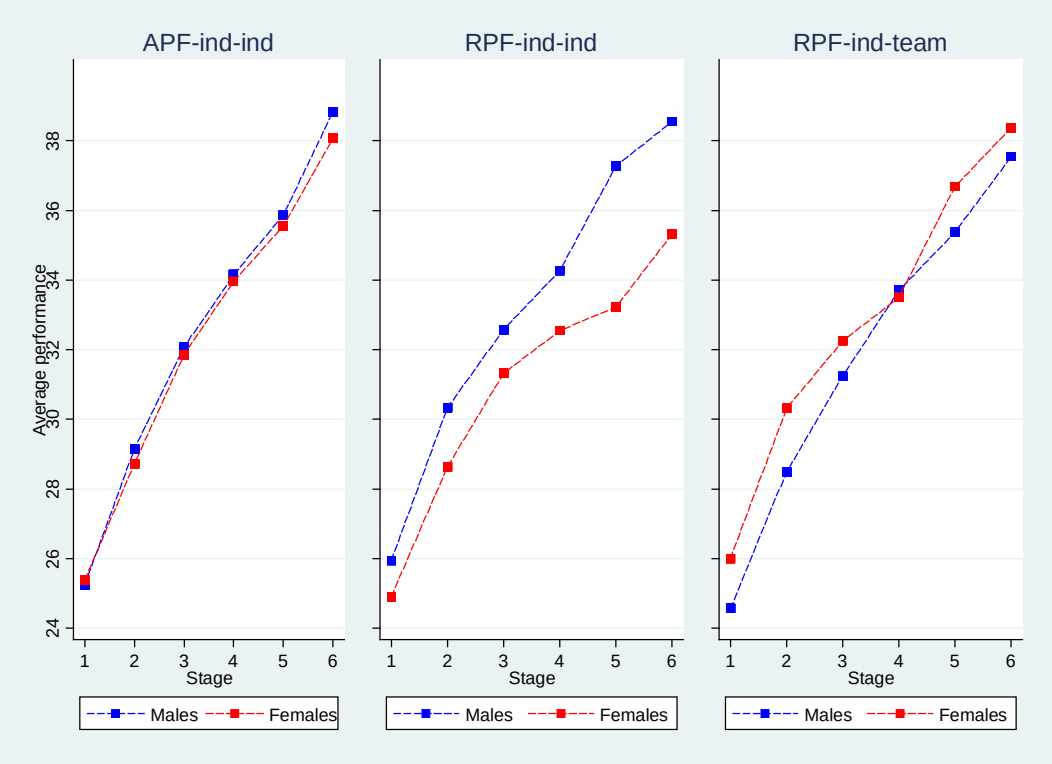


Figure A-3: Gender – team incentive treatments

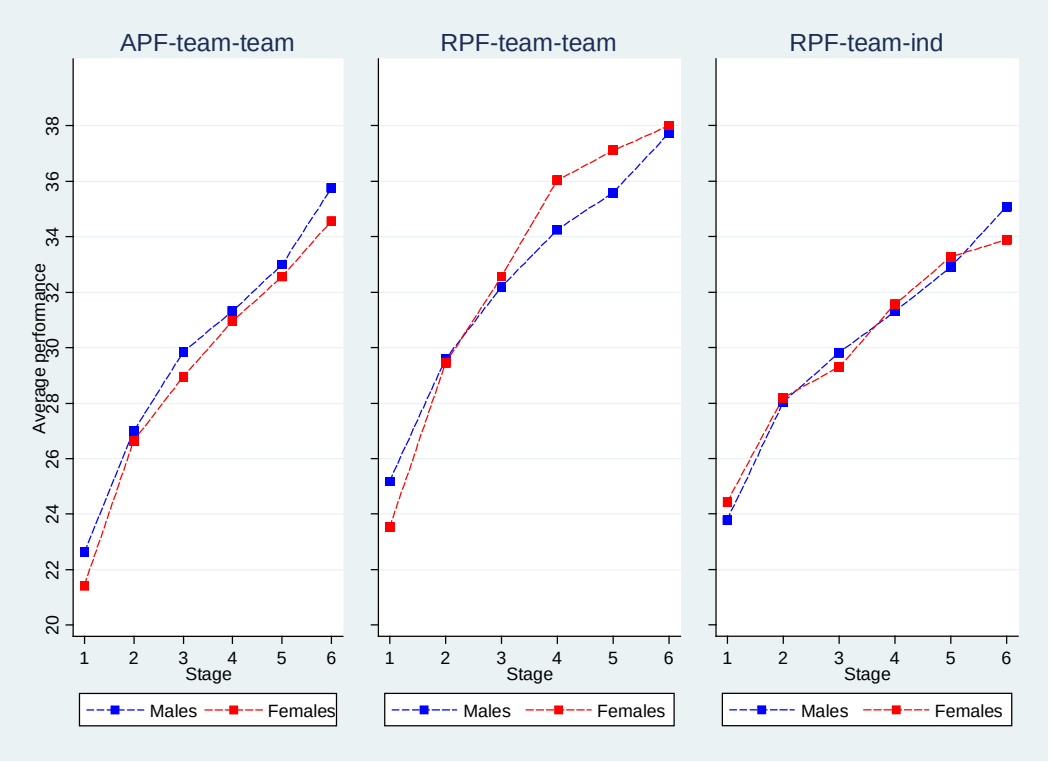
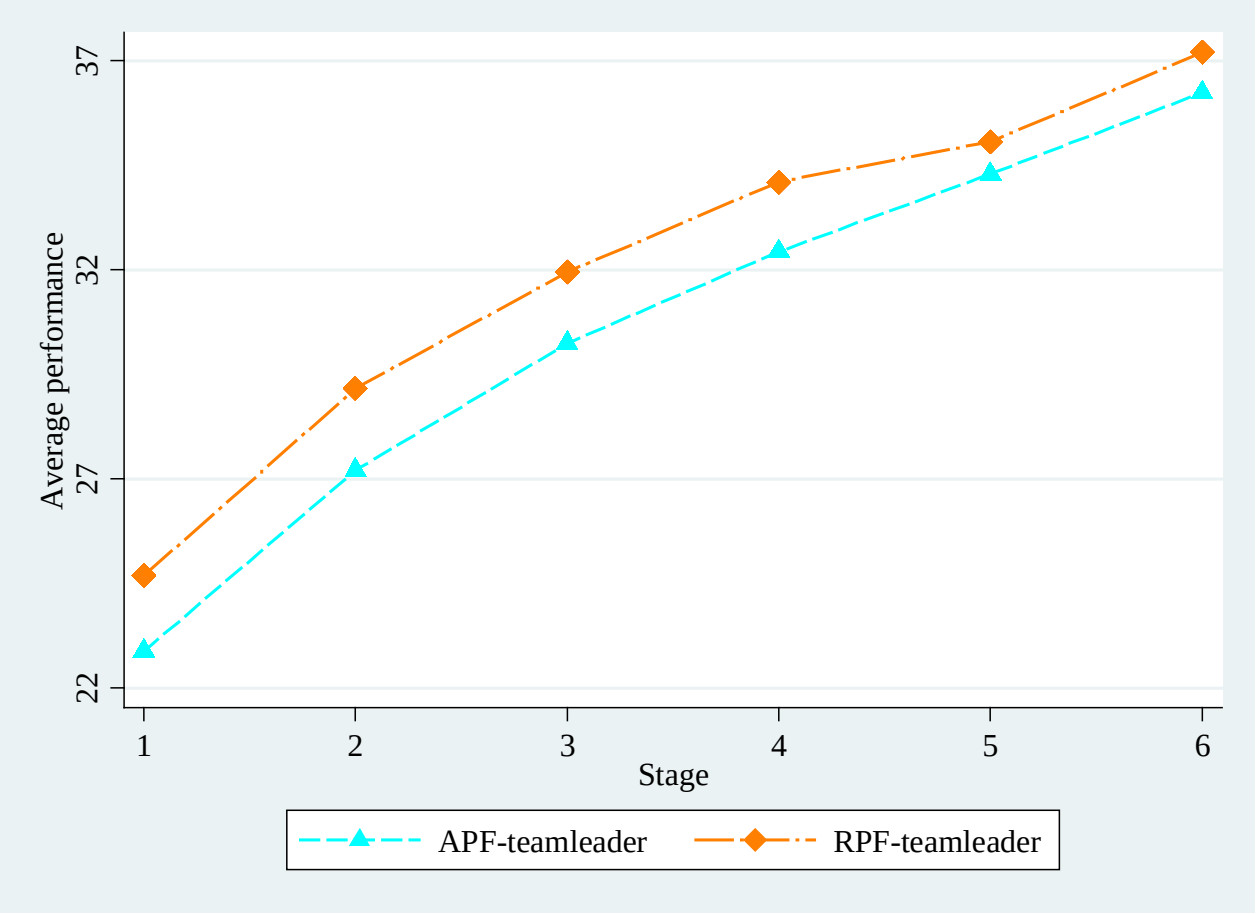


Figure A-4: Average performance across stages in teamleader treatments



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 minutes.

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 minutes.

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 minutes.

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 minutes.

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 minutes.

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 minutes.

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 minutes.

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 minutes.

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.