



Centre National
de la Recherche
Scientifique

GATE
Groupe d'Analyse et de Théorie
Économique
UMR 5824 du CNRS



DOCUMENTS DE TRAVAIL - WORKING PAPERS

W.P. 08-12

Feedback and Incentives : Experimental Evidence

Tor Eriksson, Anders Poulsen, Marie-Claire Villeval

Avril 2008

GATE Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS
93 chemin des Mouilles – 69130 Écully – France
B.P. 167 – 69131 Écully Cedex
Tél. +33 (0)4 72 86 60 60 – Fax +33 (0)4 72 86 60 90
Messagerie électronique gate@gate.cnrs.fr
Serveur Web : www.gate.cnrs.fr

FEEDBACK AND INCENTIVES: EXPERIMENTAL EVIDENCE

Tor Eriksson^a Anders Poulsen^b Marie Claire Villeval^c

8 April 2008

Abstract: This paper experimentally investigates the impact of different pay and relative performance information policies on employee effort. We explore three information policies: No feedback about relative performance, feedback given halfway through the production period, and continuously updated feedback. The pay schemes are a piece rate payment scheme and a winner-takes-all tournament. We find that, regardless of the pay scheme used, feedback does not improve performance. There are no significant peer effects in the piece-rate pay scheme. In contrast, in the tournament scheme we find some evidence of positive peer effects since the underdogs almost never quit the competition even when lagging significantly behind, and frontrunners do not slack off. Moreover, in both pay schemes information feedback reduces the quality of the low performers' work.

JEL Classification : C70, J16, J24, M52, J33, J31, C91

Keywords: Performance pay, tournament, piece rate, peer effects, information, feedback, evaluation, experiment.

^a Aarhus School of Business, Aarhus University, Department of Economics and Center for Corporate Performance, Prismet, Silkeborgvej 2, 8000 Aarhus C, Denmark. Phone +45 8948 6404, Fax +45 8948 6197, E-mail: tor@asb.dk

^b School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom. E-mail: a.poulsen@uea.ac.uk.

^c University of Lyon, Lyon, F-69003, France; CNRS, UMR 5824, GATE, Ecully, F-69130, France; Institute for the Study of Labour (IZA), Bonn, Germany. GATE: 93, Chemin des Mouilles 69130 Ecully, France. E-mail: villeval@gate.cnrs.fr.

The authors are grateful to participants at the Workshop on the Economics of the Workplace at the Tinbergen Institute, the 2nd European Workshop on Experimental and Behavioral Economics in Lyon, the ANR workshop on Incentives and Motivation in Paris, and the Conference on Tournaments, Contests and Relative Performance Evaluation in Raleigh for their comments. Financial support from the EMIR program of the French National Agency for Research (ANR) and from the Danish Social Science Research Council is gratefully acknowledged.

FEEDBACK ET INCITATIONS: UNE ETUDE EXPERIMENTALE

Tor Eriksson Anders Poulsen Marie Claire Villeval

Résumé: Cet article analyse l'impact sur l'effort des salariés de différents modes de rémunération combinés à diverses politiques d'information sur la performance relative. Trois politiques d'information sont examinées: une absence d'information sur la performance relative, une information donnée à mi-parcours au cours de la période de production, et une information continue. Les modes de rémunération à la performance considérés sont un salaire au rendement et un tournoi dans lequel le gagnant remporte l'intégralité du prix. Nous observons que quelque soit le mode de rémunération utilisé, l'information sur la performance relative n'améliore jamais les résultats des salariés. Nous n'identifions pas d'effets de pairs associés au salaire au rendement. Par contre, dans un tournoi, il existe des effets de pairs positifs dans la mesure où les concurrents dominés n'abandonnent jamais la compétition même lorsqu'ils sont largement à la traîne du premier concurrent, et les concurrents en tête ne réduisent pas leur effort. De plus, sous les deux modes de rémunération, l'information sur la performance relative réduit la qualité du travail du salariés moins performants.

Codes JEL: C70, J16, J24, M52, J33, J31, C91

Mots-clés: Salaire à la performance, tournois, salaire au rendement, effets de pairs, information, feedback, évaluation, expérience de laboratoire.

1. INTRODUCTION

An important aspect of the strategy of most organizations is the provision of incentives to the employees to meet the organization's objectives. Typically this implies tying performance to pay in the best possible manner; for a survey of the extensive literature, see Prendergast (1999). In order to reward employees for their effort, firms and organizations spend considerable resources on performance evaluations. In many cases, evaluation consists of comparing actual performance to a predefined individual target. Another frequently used format is relative performance evaluation. An extreme example is forced curve rating where managers rate each employee relative to co-workers: for example, General Electric used forced rating to eliminate 10% of its bottom employees each year.

Relative performance evaluation may motivate employees to work harder and therefore to improve the company's performance. But it may also be demoralizing and create an excessively competitive workplace, which may hinder overall performance; see Lazear (1989). While determining the overall impact of relative performance evaluation is crucial for companies, in most economic research on incentives and performance, relative performance evaluations have received only little attention. The aim of this study is add to this literature.

We address two issues. For a given pay scheme, how intensively should the organization provide its employees with information about their relative performance, when the organization's objective is to maximize the employees' total output? How does the answer to this question depend on the pay scheme used by the organization? We examine these questions experimentally by studying employee-subjects' performance under various payment schemes, namely a piece rate and a tournament scheme, and under various relative performance information policies, namely no information, information given half-way during the production period, and continuous feedback about relative performance. Our interest lies in studying performance under these various combinations of pay and information scheme. Which combination is the most profitable?

It is not hard to see why pay and relative performance information are crucial aspects of organizational design.

First, information about fellow employees' performance is likely to matter when the organization uses pay schemes that are based on *relative* performance, as in tournaments; see Lazear and Rosen (1981). In such competitive settings the organization must decide whether to inform employees about their relative standing during the competition, or to let them remain ignorant. The choice is likely to matter. An employee who is informed he is an underdog may be discouraged and lower his performance, or even quit the job. And a frontrunner who learns that he is well ahead of the other contestants may think that he can afford to slack off. In both cases, performance declines. But it is also possible that an informed underdog works harder than when uninformed to preserve pride and avoid shame or that an informed frontrunner becomes more enthusiastic and maintains or even increases effort. In this case releasing information about relative performance leads to better overall performance. Only a few theoretical models have analyzed these issues and they usually conclude that it is better to conceal information (see e.g., Lizzeri, Meyer and Persico, 2002, on the impact of feedback on own performance on effort, and Ederer, 2004, on the impact of feedback on relative performance).¹ The empirical literature about the relevance and magnitude of the effects is scant and the current paper is one of the first attempts to provide some empirical evidence on the impact of feedback on performance in the context of performance-pay.²

¹ See Goltsman and Mukherjee (2006) on the differentiation of announcements according to the agents' difference in ability, Yildirim (2005) on optimal interim revelation of ability by the agents themselves, Ertac (2005) on the information derived by the agents from the feedback given about the other agents, and Gershkov and Perry (2006) on the optimal aggregation rule between the midterm and final reviews. Another strand of the literature focuses on biased feedback. An early treatment of whether an organization can benefit for distorting information is Gibbs (1991); see Gneezy (2005) and Gürtler and Harbring (2007) for experimental evidence.

² The management literature, see e.g., Milkovich and Newman (1996), has a clear message: revealing true performance information to the employee is bad. The standard explanation, based on psychology and politics, is that the information creates animosity among employees and hurts morale.

Second, information about relative performance has been found to influence behavior even when conventional economic analysis says it should not matter. Two recent experiments, Falk and Ichino (2006) and Mas and Moretti (2006), show that allowing people to observe each other's effort has a positive influence on an employee's performance although her payment is independent of the other employees' performance. This is attributed to positive peer effects (see Kandel and Lazear, 1992), according to which employees adjust their effort to the co-worker's effort, due to non-monetary reasons such as social pressure, fear of informal sanctions, and shame. Social psychologists have also documented many non-monetary effects of information about relative performance; see Kluger and Denisi (1996). Economists have recently started to formally model how peer effects interact with monetary incentives in the workplace; see, for example, Kräkel (2007).

What is the relative role of pay and relative performance information for performance and profits? How, if at all, do pay and relative performance information interact? What is the overall effect on performance and profit of providing information on relative performance, and, if such information makes a difference, how can it be decomposed into the separate impacts on underdogs and on frontrunners in a tournament? Are the positive peer effects that have been experimentally documented for performance-independent pay schemes also present in performance-pay schemes (piece rates and tournaments)? Or, are the peer effects crowded out by the stronger monetary incentives? We define peer effects as the consequences of observing others' performance on own effort that are due to non-monetary reasons.

Our paper sheds empirical light on all these issues by means of a laboratory experiment since it is difficult to get suitable data from real organizations³. We find that on the whole, feedback on relative performance, regardless of the performance-pay scheme used, does not improve performance. Indeed, neither discrete nor continuous feedback on relative performance generates positive peer effects in the piece-rate pay scheme. In contrast, we do

³ Notable expectations are Bandiera, Barankay, and Rasul, 2005, Lazear, 1992, and Mas and Moretti, 2006.

find evidence of positive peer effects in the tournaments: frontrunners do not slack off and the underdogs almost never quit the competition, even when lagging significantly behind. These effects are however not sufficient to improve overall performance when compared to a no information environment. In addition, in both pay schemes we find what we call a negative ‘quality peer effect’: information about relative performance has a negative effect on scores because it causes the less able performers to make more mistakes and not to work less hard.

The rest of the paper is organized as follows. Section 2 described the recent literature on feedback. Section 3 describes the experimental design and delivers predictions. The data are analyzed in Section 4 and Section 5 discusses our results and concludes.

2. RELATED LITERATURE

The studies by Falk and Ichino (2006) and Mas and Moretti (2006) examine the impact of information about relative performance on effort in settings with a flat wage pay scheme. Both studies find significant positive peer effects. In the first study subjects stuff envelopes, in exchange for a fixed payment that does not depend on output. Subjects either work alone or work in pairs where each subject can see the other’s output. Falk and Ichino observe that output is considerably higher in the latter case and interpret this as evidence of positive peer effects. Their data are consistent with a model where workers experience disutility when they are falling behind their fellow workers.

Mas and Moretti (2006) examine peer effects in settings where workers (super market cashiers) are affected by other workers’ performance through externalities. Each worker works and is paid independently of each other, but has an incentive to free ride by unloading the workload on to the other workers. Sufficiently strong peer effects may, however, outweigh this incentive. The data show that this is indeed the case: More productive workers induce an increase in productivity of fellow employees, but only significantly so when the workers can observe each other’s output. This is consistent with peer effects being social

pressure (including mutual monitoring or informal sanctions). Social pressure enhances the performance of low productivity workers while it does not decrease the performance of high productivity workers. In such a context, providing employees with information about their relative performance is beneficial for the organization since it triggers desirable peer effects.

The studies by Falk and Ichino and Mas and Moretti use a flat wage pay scheme. In contrast, our paper empirically examines whether or not peer effects are present when the organization uses performance-pay (a piece rate or a tournament). The few studies that have considered the influence of competitors' ability on own effort have produced mixed evidence.⁴ A related field study is Bandiera, Barankay, and Rasul (2005). They show that when a relative pay scheme is used, a worker's productivity is affected by her co-workers' productivity as the negative externality is internalized, but only when mutual monitoring is possible. In contrast with the three previous studies, in our game the players cannot observe each other, they only observe the performance of others.

A recent tournament experiment is Ederer and Fehr (2007). In a two-period set-up subject-principals can at no cost send potentially incorrect messages about the relative performance to each worker at the end of period 1, after which workers decide on their period 2 effort. Ederer and Fehr also consider a treatment where information is truthfully revealed and one where no information is released. If marginal cost is linear (convex) [concave] then period 2 effort under truthful revelation is the same as (lower than) [higher than] under no information. Our experiment has some similarities with their study but it also differs in a number of important respects. Like them, we compare a tournament setting where no information is released with one where interim (and truthful) information is made available. In addition, we study the case where information is released continuously. This allows us to identify the potential influence of the frequency of feedback on behavior. Moreover, we use a real effort

⁴ See also Haragushi and Wadell (2007), Guryan, Kroft, and Notowidigdo (2007), and Ludwig and Ruchala (2006).

experiment, that is our subjects perform a task in real time, whereas the subjects in Ederer and Fehr (2007) make hypothetical effort choices. By letting subjects produce output in real time we are able to address some questions that a hypothetical choice experiment cannot. In our Continuous Feedback treatment, where workers are constantly kept updated on relative performance, we can directly observe the dynamics of competition. We can, for example, examine whether underdogs try to catch up or become discouraged, and if frontrunners strategically lower their work effort in response to the relative performance information. In addition, in our experiment the information that is provided is unbiased.

As already mentioned, one purpose of this paper is to examine empirically how prevalent quitting or drop out decisions are in tournaments when information about relative performance is provided, and to gauge their impact on the frontrunner's subsequent work decision. Some papers have investigated tournaments where one contestant is exogenously favored over another player (see e.g., O'Keefe et. al., 1984, Schotter and Weigelt, 1992, and Kräkel, 2007). They find that the disadvantaged (and sometimes the advantaged) contestant does not frequently drop out and chooses a larger effort than predicted by theory. In contrast, Müller and Schotter (2007) find that low-ability workers tend to drop out when promotion opportunities are limited. We do not exogenously impose differences between contestants. Instead, such differences emerge endogenously during the competitive process, as players exert effort and observe their relative positions. We feel that using such a dynamic real-effort experiment provides a better framework for studying drop out decisions.

Fershtman and Gneezy (2004) study quitting behavior by conducting a real-effort field experiment with high school students running athletics races. The students first run alone without any rewards. In the second stage students run in pairs and compete for a prize. Students were either randomly matched or matched on the basis of ability observed in the first round. Also, the monetary stakes were manipulated (none, low, high) and the tournaments

were either direct (runners ran side by side) or indirect (subjects ran separately). The latter distinction is similar to our No Feedback and Continuous Feedback treatments. Fershtman and Gneezy find that average performance is higher when matching is based on ability. Also, higher monetary stakes yield higher average performance. However, they do not find that it matters significantly for performance whether the tournament is direct or indirect. But there is significantly less quitting in the indirect than in the direct tournament. Notably, dropping out of the tournament is not very common and occurs only when incentives are strong. Fershtman and Gneezy attribute the low drop out rate to negative feelings such as shame and social stigma that such a decision may generate due to the presence of an audience. Our paper helps in studying whether dropping-out behavior is more frequent when subjects receive a feedback on others' performance but cannot observe or be observed by anyone.

3. EXPERIMENTAL DESIGN

3.1. Procedures

Our experiment uses a real task that consists of adding sets of four randomly generated two-digit numbers for 20 minutes. All the subjects in all sessions receive the same numbers in the same order. This task, also used in Niederle and Vesterlund (2007), is simple, gender-neutral, requires effort (mental concentration) but not learning, and the outcome is easy to measure. The subjects see the numbers on the computer screen and enter the answer using the keyboard. Use of a pen, scratch paper, or calculator is not allowed. Once an answer has been submitted, a new set of numbers appears on the screen. Each subject can work at his own pace. The screen displays the number of correct answers and the remaining time available. The score equals the number of correct answers.

If a subject in a tournament feels he is certain to lose he may decide that it is not worth to continue performing the task. But simply sitting still in the lab may not be a very attractive alternative either. We made a newspaper available on each desk and informed the subjects

that they were free to read this newspaper, or any book or magazine that they had brought with them. A subject could thus quit the competition by reading newspapers or books during the remaining time of the experiment, Or she could decide to take a ‘break’ for some time, and then start to work again (incidentally, this never happened).

Pay schemes

We randomly matched subjects in pairs. In the piece rate payment scheme a subject’s earnings only depend on the subject’s own score (number of correct answers). Each correct answer pays 1 point (10 points=€1). Wrong answers do not subtract points. In the tournament payment scheme a subject’s earnings depend on the relative performance. The subject with the highest score receives a fixed winner’s prize of €7 and the other subject receives €0. In case of ties the winner is randomly selected and the other subject receives nothing.

Information conditions

In all our information conditions subjects always know their own current score (current total correct answers). What they know about the other subject’s current score differs across treatments. In the No Feedback condition a subject never observes the other subject’s performance. In the Discrete Feedback condition each subject is informed about the other subject’s score halfway during the production period, that is when 10 minutes have elapsed; after that no further information is released before the end of the game. In the Continuous Feedback condition each subject can at any time observe the other subject’s score on the screen. In all three information conditions a subject observes the other subject’s score after the 20 minutes have elapsed and in the tournament conditions he is told if he won or lost. In each feedback condition the above is common information.

Appendix 1 contains a screen shot. All in all, we have $2 \times 3 = 6$ treatments, of the form (i,j) , where $i=PR$ (Piece Rate), T (Tournament), and $j=NF$ (No Feedback), DF (Discrete Feedback), CF (Continuous Feedback).

Choosing piece rate and tournament parameters

To ensure a non-arbitrary comparison of our payment schemes we chose the parameters of the piece rate and tournament (per unit payment and prizes) such that profits depend only on output (total scores). To do this we first fixed the piece rate, b , and ran the Piece Rate - No Feedback treatment. The average observed output from this treatment, denoted \bar{y} , determines an average wage, $b\bar{y}$, which was approximately equal to €8.5. We used the same average wage cost in the tournament, by setting the loser's prize equal to zero and setting the winner's prize equal to $2 \times €8.5 = €17$. This method equalizes the wage cost per worker-pair in all the tournament conditions and in the piece rate treatment with no feedback. This, in turn, implies that profitability differences exactly reflect differences in scores.

Experimental procedure

The experiment was conducted at the GATE laboratory in Lyon, France, using the REGATE software (Zeiliger, 2000). 208 undergraduate students from local business and engineering schools (101 males and 107 females) participated in 8 sessions. Each subject participated in a single treatment only. Table 1 summarizes the treatments.

Table 1. The treatments

Treatment	Payments	Information about co-player's score	# of sessions	# of subjects
Piece Rate – No Feedback (PR-NF)	€0.1 per unit output	After 20'	1	28
Tournament – No Feedback (T-NF)	€17 or €0	After 20'	2	44
Piece Rate – Discrete Feedback (PR-DF)	€0.1 per unit output	After 10' and 20'	1	30
Tournament – Discrete Feedback (T-DF)	€17 or €0	After 10' and 20'	1	28
Piece Rate – Continuous Feedback (PR-CF)	€0.1 per unit output	Throughout	1	30
Tournament – Continuous Feedback (T-CF)	€17 or €0	Throughout	2	48

Note: 'output' denotes a set of correctly added numbers.

Upon arrival the subjects were randomly assigned to a computer. The instructions, describing the nature of the task, the money payoffs and the nature of information (see Appendices 2 and 3), were distributed and read aloud. The subjects completed a

questionnaire in order to check their understanding, and practiced the task for three minutes for free. Next, each subject was randomly matched with another subject and performed the task for a maximum of 20 minutes, with or without information about the co-subject's score, depending on the information condition. A session lasted on average 75 minutes. Subjects received a show-up fee of €7. On average the subjects earned €15.41 (min=€7; max=€24). Subjects were paid in cash in a separate room.

3.2. Predictions

In this section we formulate some predictions about the behavior we expect to see in our treatments. Our objective is not to set up and test a theoretical model: Any such attempt would ultimately be unsuccessful since in our real effort experiment we do not have full experimental control over the parameters which according to theory are important, namely the effort cost function. In addition, as pointed out by Fershtman and Gneezy (2005), equilibrium behavior in the environment of a continuous time tournament has not yet been given an analytical characterization. Our experiment empirically elucidates how individuals react to interim feedback in a dynamic environment.

Peer effects

By peer effects we understand people's effort adjustment in response to other agents' scores that is due to non-monetary considerations such as social comparisons. One possibility is that people compare their score with reference agents, and that it is subjectively costly to be out of line with them (Kandel and Lazear, 1992; Kräkel, 2007). As Falk and Ichino (2006) we think of peer effects as being based on score comparisons that affect the cost of effort: A person suffers a disutility (gains in utility) from performing worse (better) than her colleagues.

Peer effects are positive (negative) if a person's marginal cost of effort falls (increases) when the other agent increases his score. *Ceteris paribus*, this leads the agent to increase (reduce) his effort. Intuitively, positive peer effects mean that lagging behind others is

unpleasant, and in order to avoid experiencing this feeling the agent seeks to catch up with the other agent. Similarly, outperforming others is pleasant, and the agent seeks to increase his distance with the other agent. If peer effects are negative, the agent becomes discouraged and further reduces effort, or even quits although it is rational in terms of payoff maximization to continue exerting effort.

Recall that in our experiment effort equals submitted answers, but if an answer is false no output is produced: only correct answers generate scores. In other words, effort (answers) = score (correct answers) + mistakes (incorrect answers). This has the fundamental implication that an increase in effort may not be accompanied by a similar increase in scores if the number of mistakes increases. The important distinction between scores and effort means that in our set-up there is another potential peer effect, which does not affect the quantity of effort but rather its ‘quality’, namely the extent to which effort is translated into score. In the rest of the paper, ‘*peer effect*’ refers to the peer effect operating on the quantity of effort, and we refer to the one operating on quality as a ‘*quality peer effect*’. The quality peer effect is negative (positive) when an increase in the other agent’s score leads the agent to make relatively more (fewer) mistakes. A positive quality peer effect means that the agent makes relatively fewer mistakes. We expect quality peer effects to be predominantly negative.

The Piece Rate pay scheme

There is no strategic interaction between the agents under the piece rate scheme, so if there are no peer effects each agent’s behavior will be the same regardless of the information feedback policy.

Hypothesis 1: Suppose there are no peer effects. Then the effort and the score distributions in the PR-NF, PR-DF, and PR-CF treatments are the same.

The presence of peer effects qualitatively transforms the piece rate work environment since agents’ subjective welfare becomes interdependent: If agent i increases his effort, there is an

externality on agent j 's welfare, since agent j 's cost of effort is influenced by the change in agent i 's output, and this leads agent j to revise his effort decision, which in turn has a feedback on agent i , and so on.

Positive (negative) peer effects in the piece rate means that a person's effort choice is increasing (decreasing) in the other person's output. Obviously, this requires that the agent must be informed about his co-participant's score over time. Consider the Piece Rate – Discrete Feedback treatment. The two agents produce a score in Period 1 (0 – 10 minutes), and at the end of the period both agents learn the Period 1 score difference. It is quite straightforward to verify that if there are positive (negative) peer effects, an agent's Period 2 effort will be increasing (decreasing) in the other agent's Period 1 score. This gives the following proposition:

Proposition 1: Consider the Piece Rate – Discrete Feedback treatment. If there are positive (negative) peer effects, an agent's period 2 effort will be increasing (decreasing) in the other player's period 1 score.

We expect peer effects, if any, to be strongest in the Piece Rate with Continuous Feedback. We also expect the within-pair scores to be closer (more distant) in the presence of positive (negative) peer effects. The reason is that in the presence of positive peer effects, the agent who is lagging behind tends to catch up and both effort levels tend to become more similar than when there are no peer effects.

Hypothesis 2: If there are positive (negative) peer effects, total effort and scores in the Piece Rate treatments will be ranked as follows: PR–CF>PR–DF>PR–NF, and the within-pair differences will be ranked as follows: PR–NF>PR–DF> PR–CF.

The Tournament pay scheme

The tournament differs fundamentally from the other schemes in that the monetary payment to an agent depends on the other agent's performance, and as a consequence there is strategic interaction between the agents even in the absence of peer effects. Peer effects, whether affecting effort or quality of effort, introduce an additional strategic component which may interact with the one generated by the monetary earnings. We first describe the strategic effects predicted by money maximization, and then turn to the impact of peer effects.

Ederer and Fehr (2007) compare a tournament with no information with the case where agents receive feedback about relative performance halfway through the experiment. Thus, their model matches our experimental No Feedback and Discrete Feedback treatments, except for the fact that in our experiment we do not introduce stochastics since this could make the task less transparent for the subjects. Nevertheless, we use the predictions by Ederer and Fehr (2007) as a benchmark. In what follows we denote the person who produced most (least) during Period 1 as the frontrunner (underdog).

Proposition 2: In the T-DF treatment and in the absence of peer effects both the frontrunner and the underdog's Period 2 scores are decreasing in the Period 1 score difference.⁵

This proposition implies that absent of peer effects, providing a more intense feedback on relative performance has a negative overall effect on effort. The intuition is the following. Feedback about relative performance reduces the underdog's effort since he revises his beliefs about the chance of winning downwards and could even lead him to drop out altogether. Also, we predict that information feedback tends to reduce the frontrunner's effort incentives: on learning that he is ahead, he revises upwards his belief of winning which can lead him to reduce his intensity of effort.

Hypothesis 3: Suppose there are no peer effects. Then total effort in the tournament

⁵ Furthermore, following Ederer and Fehr (2007), if marginal cost of effort is linear (convex) [concave], then total output in T-NF is the same as (smaller than) [bigger than] than total output in the T-DF treatment. While this result is of significant theoretical interest, it provides however little guidance for our data analysis since we do not control our subjects' cost function.

treatments is predicted to rank as follows: $T\text{-CF} < T\text{-DF} < T\text{-NF}$. Moreover, in the Tournament - Discrete Feedback the Period 2 effort of underdogs with similar Period 1 effort will be decreasing in the Period 1 output gap.

Consider next the impact of peer effects. Suppose peer effects are positive, meaning as before that an agents' marginal cost of effort is decreasing in the other agent's score. In the T-DF or T-CF treatment, when an underdog learns he is lagging behind the frontrunner, the peer effect leads him to increase his effort. This tends to counteract the negative effect on the underdog's effort due to the monetary incentives. Of course an increase in effort may not translate into increased score on a one-to-one basis, if accompanied with a negative quality peer effect (some effort is then "lost" through mistakes). Consider then the frontrunner: Being informed that he is ahead, a positive peer effect will induce him to increase or at least maintain effort. This offsets the monetary incentive to reduce effort.

This argument implies that when positive peer effects are present we should expect a more intense feedback policy to increase overall effort levels relative to a policy which provides less intense feedback about relative performance. This peer effect may of course not be strong enough to change the ranking predicted by Hypothesis 3. Thus, we can only predict the opposite effort ranking, $T\text{-CF} > T\text{-DF} > T\text{-NF}$, in the case where the positive peer effects are so strong that they dominate the negative impact of the underlying monetary incentives. Consequently, we can not give the 'converse' of Hypothesis 3 but have to be content with a partial version:

Hypothesis 4: Assume there are positive peer effects. Then, in T-DF, the period 2 effort of underdogs with similar Period 1 effort are increasing in the Period 1 score gap.

This hypothesis is analogous to Hypothesis 2 for the Piece Rate. Positive peer effects are

predicted to induce the underdog to try to catch up.⁶

Quality peer effects

In both payment schemes it is possible that a positive peer effect is (partially) outweighed by a negative quality peer effect. For example, an agent who lags behind and who is driven by comparisons to work harder can at the same time become stressed and anxious, and begin to make more mistakes, in which case some of the extra effort is wasted.⁷ In order to detect a quality peer effect in the data we must distinguish it from the effort peer effect. We propose the following test: Suppose agents' effort, e , is the same in any two treatments but the proportion of mistakes m in the treatment with more intense feedback is larger at the expense of the proportion of scores s (i.e., m/e is larger and so s/e smaller). This provides evidence that increasing feedback about relative performance generates a negative quality peer effect. This gives us the following hypothesis:

Hypothesis 5: If there are negative quality peer effects the proportion of mistakes made by the agents who are lagging behind in both payment schemes will be the highest in the Continuous Feedback treatments and lowest in the No-Feedback treatments.

4. RESULTS

We first study the impact of the pay scheme and information conditions on aggregate performance and on the within-pair differences in performance. We then investigate the presence of peer effects. Finally, we consider quitting decisions.

⁶ Note that we cannot as easily develop a strategy for identifying negative peer effects. These are predicted to lead the underdog to reduce effort, which is the same that a hypothesis based on no peer effects predicts. It suggests that an underdog may decide to quit although he still had a chance to win.

⁷ An extreme example of this is the so-called 'choking under pressure' phenomenon (Baumeister, 1984), where the agents' performance deteriorates when the performance act itself becomes more important.

4.1. Feedback, incentives, and aggregate performance

Table 2 shows average scores (number of correct answers), average mistakes (number of incorrect answers), and average submissions (the sum of scores and mistakes), and their standard deviation.

Table 2. Average and standard deviation of scores, mistakes, and submissions

Treatments	Score		Mistakes		Submissions	
	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
PR-NF (28)	84.00	20.79	10.86	7.63	94.86	19.52
PR-DF (30)	83.97	24.62	12.10	11.24	96.07	26.59
PR-CF (30)	81.50	24.40	13.30	7.25	94.80	20.88
Average (88)	83.14	23.15	12.11	8.87	95.25	22.35
T-NF (44)	80.30	17.12	9.59	6.28	89.89	15.02
T-DF (28)	75.93	25.91	11.36	5.84	87.29	26.03
T-CF (48)	76.33	23.04	12.17	6.61	88.50	21.67
Average (120)	77.69	21.72	11.03	6.37	88.72	20.54

Note: The number of observations by treatment is given in parentheses.

The average score in the Piece Rate condition do not differ significantly from those in the No Feedback and in the Discrete Feedback treatments ($p=0.970$). The average score is somewhat lower in the Continuous Feedback treatment, but not significantly so ($p=0.558$).⁸ On average, subjects solve 84 problems in PR-NF and in PR-DF, while they only solve 81.5 problems in PR-CF. The difference in the average number of submissions (i.e., the level of effort) is not significantly different between the No Feedback condition and the Discrete Feedback condition ($p=0.939$) or the Continuous Feedback condition ($p=0.858$).

In the Tournament pay condition we find no significant differences in average score across the three information conditions. The average score is smaller in the Continuous Feedback treatment (76.33) than in the No Feedback treatment (80.30), but the difference is not significant ($p=0.190$). The difference between the average score in the Discrete Feedback treatment (75.93) and the No Feedback treatment is also not significant ($p=0.230$). The

⁸ In the non-parametric statistics reported here, each subject in the PR-NF and T-NF treatments gives one independent observation, whereas each pair gives one independent observation in all the other treatments. Unless explicitly mentioned otherwise, these tests are Mann-Whitney rank-sum tests.

difference in the average number of submissions is not significantly different either between T-NF and T-DF ($p=0.462$) or between T-NF and T-CF ($p=0.564$).

Result 1: Releasing information about relative scores does not significantly influence the average final score or the average effort in neither the Piece Rate or the Tournament pay condition.

This result supports Hypothesis 1 against Hypothesis 2 and rejects Hypothesis 3, at the aggregate level.

Are there, for a given information condition, significant score differences between the Piece Rate and the Tournament schemes? Table 2 shows that for any information condition, average scores drop when there is a switch from Piece Rates to the Tournament. For the No (Discrete) [Continuous] Feedback condition, the switch reduces average scores by 4.4 % (9.6 %) [6.3 %], but these differences are not significant (NF: $p=0.429$, DF: $p=0.285$, CF: $p=0.189$). The same applies to the average number of submissions (NF: $p=0.188$, DF: $p=0.256$, CF: $p=0.145$). A Kolmogorov-Smirnov test also rejects the presence of significant differences in the distribution of scores between the two schemes in the No Feedback ($p=0.271$), the Discrete Feedback ($p=0.398$), and the Continuous Feedback conditions ($p=0.219$).⁹

Is there an impact of information about relative score on the quality of effort? For the Piece Rate condition the number of mistakes (incorrect answers) under Continuous Feedback (13.30) is significantly higher than under No Feedback (10.86) ($p=0.049$). Providing discrete feedback does not significantly raise the number of mistakes ($p=0.429$). The proportion of mistakes out of total submissions increases from 11.45% in PR-NF to 12.59 % in PR-DF, and to 14.03% in PR-CF.

⁹ It should, however, be mentioned that if we pool together both conditions with interim feedback, we find significant differences between scores in the Piece Rate and the Tournament schemes: the average score is lower in the tournament (76.18) than in the piece rate scheme (82.73) ($p=0.086$). The same conclusions follow when we focus on the number of submissions that is lower in the tournament (88.05) than in the piece rate scheme (95.43) ($p=0.072$).

In the tournaments the number of mistakes under Continuous Feedback (12.17) is significantly higher than under No Feedback (9.59) ($p=0.005$). Providing feedback halfway has no significant influence ($p=0.150$). The mistake proportions in T-NF, T-DF, and T-CF are 10.67%, 13.01%, and 13.75%, respectively. This yields Result 2, which supports Hypothesis 5.

Result 2: In both the Piece Rate and the Tournament pay conditions the quality of effort is significantly lower when subjects receive continuous feedback about their relative score than when they receive no information.

Overall, regardless of whether pay depends on absolute or on relative score, an intense and ongoing comparison with co-participants' performance does not affect the quantity of effort but reduces its quality.

4.2. Feedback, Incentives, and Within-Pair Performance

Table 3 shows the average scores of the best and the worst performers in each pair, the absolute average within-pair score difference, and the average number of mistakes made by the best and the worst performers in each pair.

Table 3. Summary statistics on within-pair scores and mistakes

Treatment	Average score		Absolute within-pair score difference		Total mistakes	
	Best	Worst	Mean	St. Dev.	Best	Worst
PR-NF	94.14 (16.39)	73.86 (20.19)	20.28	16.53	11.21 (4.66)	10.50 (9.94)
PR-DF	95.20 (22.49)	72.73 (21.90)	22.47	17.96	13.93 (14.99)	10.27 (5.47)
PR-CF	95.00 (20.06)	68.00 (20.98)	27.00	16.93	10.73 (4.61)	15.87 (8.58)
Average	94.80 (19.42)	71.48 (20.72)	23.32	17.00	11.98 (9.42)	12.25 (8.40)
T-NF	87.64 (14.76)	72.95 (16.40)	14.68	14.04	8.77 (5.46)	10.41 (7.05)
T-DF	86.93 (23.78)	64.93 (23.84)	22.00	19.01	11.07 (5.88)	11.64 (6.01)
T-CF	91.42 (18.15)	61.25 (16.74)	30.17	21.16	8.63 (5.01)	15.71 (6.16)
Average	88.98 (18.29)	66.40 (18.91)	22.58	19.26	9.25 (5.39)	12.82 (6.81)

Note: Standard deviations in parentheses.

Figure 1 shows the average best–worst scores by pay scheme and information condition.

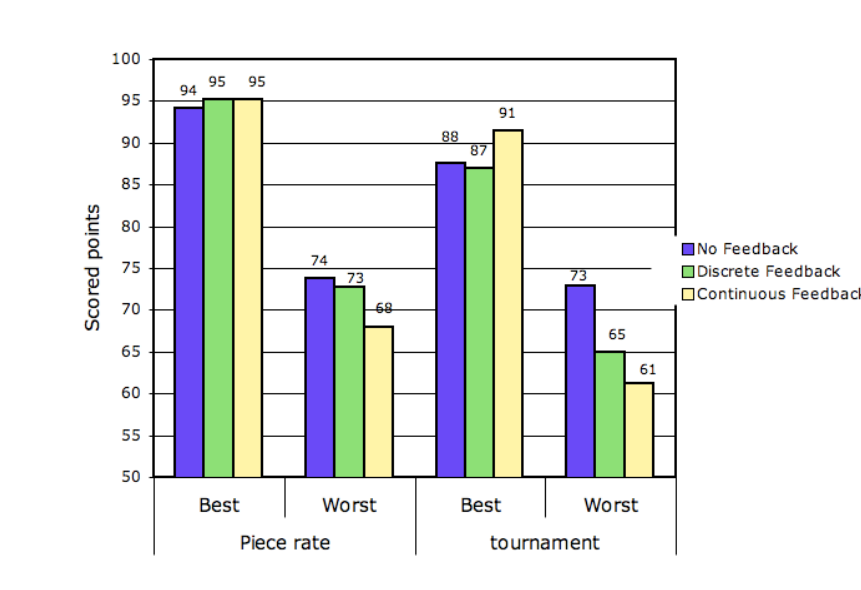


Figure 1: Average best-worst scores by pay and information condition

Table 3 and Figure 1 show that for each payment scheme the average within-pair score difference is larger when the subjects are continuously informed about their relative score than in the other information conditions. In the Piece Rate scheme with No (Discrete) [Continuous] Feedback, the average score differences are 20.28 (22.47) [27.00], respectively. However, the PR-NF does not differ significantly from PR-DF ($p=0.827$), nor from PR-CF ($p=0.336$). In the Tournament scheme, the difference between the average score differences of the Continuous Feedback and No Feedback treatments (30.17 and 14.68) is highly significant ($p=0.003$), while the difference between Discrete Feedback and No Feedback (22.00 and 14.68) is not significant ($p=0.269$).

What gives rise to the increased within-pair score differences when feedback is continuous? Table 1 shows that for each payment scheme the average score of the highest performing subjects in pairs is roughly the same across information conditions ($p=0.868$ between PR-NF and PR-CF, $p=0.628$ between T-NF and T-CF). Thus, it follows that the increasing average

score difference in the intensity of feedback is due to the lowest performing subjects reducing their scores.¹⁰ Summing up, we have:

Result 3: In the Tournament pay scheme the average within-pair final score difference is significantly higher, and the average score of the lowest performers is significantly lower ($p=0.020$) in the Continuous than in the No Feedback treatments.

This lower score of the worst performers is due either to lower effort (fewer submitted answers) and/or to an increase in the number of mistakes (incorrect answers) made by these subjects. Intense information about relative performance may discourage the subject who is lagging behind, leading to lower effort. Or, it can generate a higher level of stress which leads to more mistakes (effect observed at the aggregate level, cf. Result 2). Which is the most important factor? Figure 2 shows the average ratio of mistakes over submissions, by relative position and treatment.

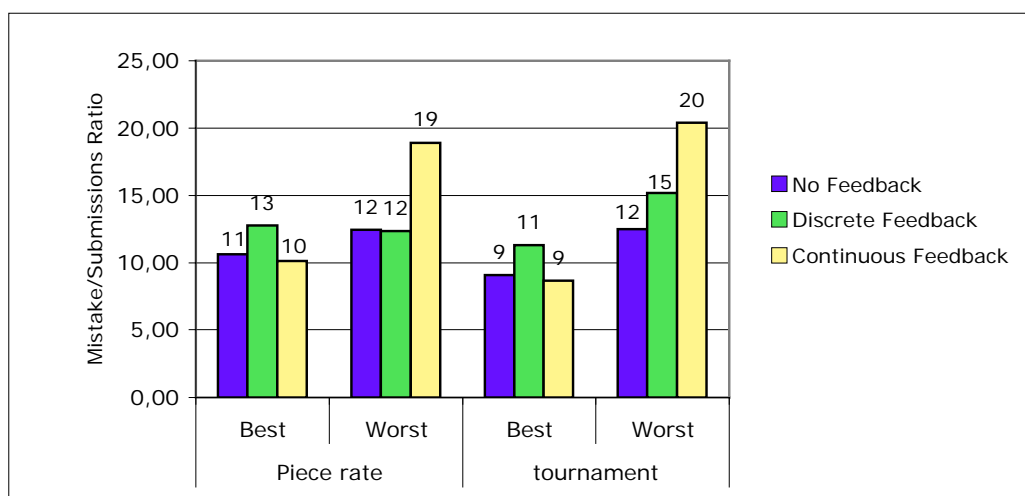


Figure 2: Average mistakes/submissions ratio by relative position and by treatment

Irrespective of the pay condition the average ratio of mistakes to submissions of the worst performers is highest in the Continuous Feedback condition (18.92 % in PR-CF and 20.41 %

¹⁰ Kolmogorov-Smirnov tests fail to find significant differences between the quantile distributions of score across treatments in the Piece-Rate condition ($p>0.10$). But in the Tournament condition, the quantile distribution of T-NF differs from both T-DF ($p=0.095$) and T-CF ($p=0.10$). This is attributable to the lowest quartile subjects who perform less well in T-CF than in T-NF. This is consistent with the observation that the standard deviations of scores in T-DF (25.91) and T-CF treatments (23.04) are higher than in T-NF (17.12).

in T-CF, compared with 12.45 % and 12.49 % in PR-NF and T-NF). In the Piece-Rate scheme the difference between NF and CF is significant ($p=0.015$), but not between NF and DF ($p=0.383$). If one considers the total number of submissions (scores + mistakes), the worst performers submit on average the same number of answers in the three information conditions (84 in PR-NF, 83 in PR-DF, and 84 in PR-CF) ($p=0.760$ between NF and DF and $p=0.776$ between NF and CF). Hence the worse scores of the lowest performers are mainly due to an increase in mistakes and not to a fall in effort (submissions).

For the tournaments we arrive at the same conclusion: the ratio of mistakes to submissions of the underdogs is significantly higher ($p<0.001$) and their score significantly lower ($p=0.020$) when continuous feedback is provided than when no information is given. Moreover, the average number of submissions by the underdogs is slightly lower when information is provided (76.57 in T-DF and 76.96 in T-CF vs. 83.36 in T-NF), but these numbers are not significantly different ($p=0.338$ between NF and DF, $p=0.183$ between NF and CF). The underdogs' scores fall because mistakes crowd out correct answers, for (roughly) same number of submissions. We can thus quite confidently conclude that increased information creates increasing score differences because the underdogs make more mistakes, and not because they work less. This yields Result 4.

Result 4: Providing continuous feedback does not significantly affect the average score of the best performers, but it reduces the average score of the worst performers, mainly due to an increase in mistakes and not to a reduction in effort. This reduction in scores is significant in the Tournament pay condition.

4.3 Peer Effects, Evolution of Effort and Quality

Recall that by 'peer effects' we mean adjustments in subjects' effort that are due to non-monetary considerations. Positive (negative) peer effects refer to a subject increasing (decreasing) her effort when she learns that the other person has produced more than herself. It is important to notice a distinguishing feature of our study. Unlike related papers in the

literature, the subjects in our experiment (Discrete and Continuous Feedback treatments) do not literally observe each other (only each other's performance) and do not know or learn who their contestants are. Thus, peer pressure does not operate via monitoring, audience effects, and alike. To understand how such effects manifest themselves in our data, we next analyse the dynamics of effort and mistakes over time.

Consider first the Piece Rate condition. Positive (negative) peer effects would imply that a subject, who in the PR-DF treatment after 10 minutes learns that he is trailing behind the other subject, exerts more (less) effort during the second period (10 to 20 minutes) *relative* to the effort of similar subjects in the same period in PR-NF. Similarly, in PR-CF a subject who is repeatedly informed that he is lagging behind should increase (reduce) his effort relative to a similar subject during the same period in PR-NF. We compute a ratio relating the number of submissions in the last 10 minutes to the number of submissions during the first 10 minutes. There is no significant difference in this ratio between PR-NF and PR-DF ($p=0.458$) or PR-CF ($p=0.419$).¹¹ This is a first indication of the absence of peer effect in the evolution of effort over time in the piece-rate payment scheme.

In the Tournament condition peer effects are more difficult to identify since they can be confounded with effort adjustments made for monetary reasons. One of our predictions for the T-DF treatment is that, in the absence of peer effects, the average effort in the second half will be decreasing in the score difference halfway through the production period. The larger this difference, the lower is both the frontrunner's and the underdog's effort. It follows that an indication of positive peer effects in T-DF is a subject who, on having been informed that he is an underdog, does not submit significantly fewer answers in the second half, compared to the submissions of similar subjects in the same period in T-NF. In T-CF, positive peer

¹¹ We acknowledge that computing data on effort after 10 minutes is arbitrary in the case of the Continuous Feedback treatment since the subjects probably adjust their effort continuously. This simplification is made here to make comparisons with the Discrete Feedback possible.

effects give rise to a similar pattern, where underdogs do not submit significantly less answers than is observed in T-NF.¹² Comparing the ratio of the number of submissions in the last 10 minutes to the number of submissions during the first 10 minutes, we find that there is no significant difference between T-NF and T-DF ($p=0.559$), whereas the difference between T-NF and T-CF is close to significance ($p=0.105$).

These non-parametric statistics do not, however, control for the similarity of effort during the first half of the period across treatments. To account for this we estimate two linear regression models. In the first, the dependent variable is the ratio of the number of submissions in the first 10 minutes to their number in the last 10 minutes. In the second, the dependent variable is the corresponding ratio for mistakes. The explanatory variables include the score after 10 minutes, the subject's gender, and dummy variables for the Discrete and Continuous Feedback treatments, the No Feedback condition being the omitted reference category. We interact these variables with the co-participant's score after 10 minutes and own gender. We hypothesize that the distance to the competitor influences the evolution of effort, and allow this effect to differ for the genders. The models are estimated separately for the underdogs and the frontrunners; see Table 4.¹³

¹² In T-CF, this effect could manifest itself earlier or later as the subjects have several opportunities to adjust their behavior. We analyze effort after 10 minutes to compare with T-DF. See footnote 11.

¹³ The results for the Piece-Rate condition are omitted in Table 4 since no coefficient was significant. This confirms the absence of peer effects on the evolution of effort exerted by the subjects under this payment scheme and provides further support for Hypothesis 1.

Table 4. Determinants of the evolution of effort in tournaments

Dependent variable	Ratio of submissions in the last 10' to submissions in the first 10'		Ratio of mistakes in the last 10' to mistakes in the first 10'	
	Underdog	Front runner	Underdog	Front runner
Score after 10 minutes	0.002 (0.002)	-0.006*** (0.001)	0.064*** (0.020)	0.028* (0.016)
Male	-0.007 (0.062)	0.041 (0.038)	0.454 (0.529)	0.330 (0.441)
Discrete Feedback treatment (DF)	-0.002 (0.162)	-0.031 (0.086)	1.265 (1.363)	1.837* (1.006)
Partner's score after 10' in DF	0.000 (0.004)	0.001 (0.003)	-0.009 (0.031)	-0.044 (0.032)
Male*Partner's score after 10' in DF	0.001 (0.002)	-0.001 (0.002)	0.000 (0.020)	-0.004 (0.021)
Continuous Feedback treatment (CF)	0.395*** (0.143)	0.002 (0.075)	2.592** (1.203)	0.514 (0.877)
Partner's score after 10' in CF	-0.008*** (0.003)	0.000 (0.003)	-0.050* (0.026)	-0.031 (0.033)
Male*Partner's score after 10' in CF	0.002 (0.002)	-0.005** (0.002)	0.010 (0.015)	-0.001 (0.021)
Constant	1.002*** (0.095)	1.322*** (0.061)	-1.116 (0.809)	0.066 (0.724)
Number of observations	60	60	58	59
Adjusted R ²	0.010	0.459	0.145	0.023

Notes: *, **, and *** indicate statistical significance at the 0.10, 0.05, and 0.01 level, respectively. Standard errors are in parentheses. We have fewer observations in the analysis of mistakes than in the analysis of submissions because 3 subjects did not make any mistake.

Consider first the underdogs. They disregard the discrete feedback, perhaps because they do not consider it as informative about their chances to win (although 13 subjects of the 14 who were behind halfway eventually lost the competition). On the other hand, controlling for their score, they increase their effort significantly more in T-CF than in T-NF, which suggests some peer effects are present. This effect is weakened by the observation of a larger distance to the competitor. This suggests that, except when a too big gap between contestants generates discouragement, continuous feedback creates emulation or positive rivalry via peer effects.¹⁴ We also find that the underdogs make significantly more mistakes in the second half of the game relative to the first half in T-CF than in T-NF. This pattern is consistent with negative quality peer effects, possibly due to stress and anxiety associated with continuous comparison of performance.

¹⁴ This result is consistent with Leonard and Van Long (2008)'s model of goal setting.

Next, consider the frontrunners. Subjects with a higher score after 10 minutes decrease their rate of submissions and increase their rate of mistakes in the second half of the game, presumably due to a fatigue effect. In the feedback treatments, we find that controlling for their score the frontrunners do not adjust their effort to the distance with the underdogs. This should, however, be qualified when feedback is continuous: when the underdog's score is higher, males' effort increase less over time than that of females, but the magnitude of this effect is very small. This could be due to different attitudes toward risk (Eckel and Grossman, 2003) or to differences in overconfidence across genders (Barber and Odean, 2001): female frontrunners need more distance to the underdogs before they feel safe to reduce their own effort. Last, a marginally significant negative quality peer effect is identified in T-DF. Recall that the peer effects observed in the DF and CF treatments arise despite the fact that, unlike in previous studies, the contestants do not observe each other. Simply observing the contestant's performance and knowing that one's own performance is observed appear to be enough to generate these effects. Overall, these findings yield Result 5:

Result 5: There is no evidence of positive peer effects on the evolution of effort in the Piece-Rate treatments. In contrast, in the Tournament treatments we find both positive peer effects and negative quality peer effects on the evolution of the underdogs' score. Positive peer effects may also be at work among frontrunners since they do not adjust their effort to the distance to the competitor.

4.4 Quitting

We define quitting as permanently stopping to work before the end of the game. There is no rationale for such behavior in the piece rate condition. But in a tournament setting, in the absence of peer effects, quitting is rational for an underdog when the likelihood of winning becomes sufficiently low. On the other hand, a person may refuse to quit even if she is sure to lose because she wishes to avoid a feeling of shame due to being known to give up. For frontrunners, it is rational to quit if the underdog has dropped out. Again, peer effects and emotions may prevent this from taking place: being ahead may generate a positive peer effect

and generate positive feelings such as enthusiasm or joy from being the best, and this may motivate the frontrunner to continue working at a high pace.

Appendices 4 and 5 show for each pair of workers the time evolution of performance in the T-DF and T-CF treatments. Although the losers of the tournaments are on average lagging behind their competitor during 92.26% of the twenty minutes, in all treatments but one all subjects work until the 20th minute. In the T-CF treatment, only 3 subjects quit before the end of the game: one person quits at 14'08" and his co-worker responds by quitting at 16'59" (see pair 4 in Appendix 5), and 17'28" for another winner (pair 16 in Appendix 5). Visually inspecting the dynamics of effort in each subject pair of subjects confirms that the subjects almost never drop out.

Result 6: In the Tournament - Discrete Feedback and Continuous Feedback treatments both underdogs and frontrunners almost never drop out, even when there seems to be little doubt about who will win.

Why do so few people quit? First, as we do not measure subject' beliefs we cannot exclude the possibility that underdogs who continued to work actually thought they might win. However, we do not find this a likely explanation: most of the performance graphs (Appendices 4 and 5) show that underdogs continue to perform the task right to the end, even though the performance gap increases. We instead attribute the lack of quitting to positive peer effects. The underdogs continue to work in order to avoid unpleasant feelings, such as shame or losing face by being seen to give up. In our experiment there is no audience (such as in Ferhstman and Gneezy, 2005), but it is still the case that the underdog's competitor learns how much the competitor produced compared to his own score at the end of the game. There may also be non-peer effects among underdogs, such as a desire to maintain a self-image as someone who never gives up and who always does one's best. Future work should seek to carefully disentangle these effects and evaluate their relative magnitude.

All in all, it is preferable to conceal information to the agents in a tournament, and this is reinforced by the presence of negative quality peer effects when such interim feedback is provided. But on the other hand, once the agents get such feedback, the tournament pay scheme benefits from the presence of positive peer effects. Underdogs do not give up, and consequently frontrunners do not slack off. Therefore, provided the quality of effort is not a major issue, if the agents are able to collect interim feedback on their relative performance, peer effects contribute to make the tournament more efficient than expected.

5. DISCUSSION AND CONCLUSION

We use a real-effort experiment to investigate the efficiency of various relative performance feedback systems under a piece rate and a tournament pay scheme. We find that releasing interim feedback about relative performance does not improve performance under either pay condition. When performance-pay is of the piece-rate type, people only try to maximize their monetary payoff and their quantity of effort is not affected by performance comparisons. On the other hand, in tournaments with continuous feedback the evolution of effort of the underdogs and the frontrunners between the first and the second half of the production period indicates the presence of positive peer effects: people adjust their effort to the performance of others less than they would do if they were only motivated by monetary considerations. As the subjects are not in visual contact, the results imply that for peer effects to arise it may be enough that the contestants know each other's performance and that you know your performance is seen by others.

Overall, the positive peer effects are not, however, sufficient to compensate for the negative effect of feedback on final performance under a competitive performance-pay scheme, and the profitability of the organization, here measured by total output, is decreased by the dissemination of information. Our conclusion is, therefore, that it is more profitable not to provide interim feedback about relative performance. This conclusion is reinforced if one

accounts for the fact that in reality providing information is usually costly for the organization.

Another important finding of our experiment is based on the observation that while continuous feedback does not lead the frontrunners to slack off in tournaments, it significantly reduces the performance of the underdogs. The negative impact of continuous feedback on underdogs is not due to a reduction in their *quantity* of effort but rather to a decrease in the *quality* of their effort. The same is observed in the piece rate scheme. We refer to this as a negative ‘quality peer effect’. Information feedback and the output comparisons they allow generate stress and anxiety that reduce the quality of effort. This is similar to how increased conscious attention to one's own process of performance, implicit competition, cash incentives, or the presence of an audience can reduce performance, as emphasized in the “choking under pressure” literature (Baumeister, 1984; Ariely et al., 2005; Dohmen, 2008). Our results indicate that when an organization also cares about the performance of the less able agents, it should not implement a continuous feedback system, whatever the performance-pay mode.

Another noteworthy finding of our study is that dropping-out from the competition is rare. Both frontrunners and underdogs continue to exert high effort even when the distance between them is so large that it seems certain that the underdog cannot catch up, that is, when the marginal benefit of effort is null. We offer several explanations for these observations. First, both frontrunners and underdogs may be motivated by competitiveness, in which case they attempt to minimize the score difference relative to their co-contestant. Second, as also discussed by Fershtman and Gneezy (2005), a social norm (‘one should never give up’, ‘always do your best’) may be at work, especially when the subject (as in Fershtman and Gneezy’s experiment) or his performance (our experiment) is observed. Third, ego (personal ambition possibly activated by the joy of winning, self-esteem or self-image) is another

possible source of motivation that could drive the subjects to attempt at maximizing their own score even when there are no more monetary incentives to exert effort. Further research is needed to distinguish between these various sources of motivation.

The existence of peer effects has been documented for a flat wage scheme. We found evidence of positive peer effects when subjects are paid under a tournament payment scheme, but not under a piece rate scheme. How can we reconcile these findings? One explanation is that peer effects are less prevalent in the piece rate than in the other pay schemes because each additional unit of effort has a direct and certain monetary counterpart that focuses attention on monetary aspects of the work environment.

Our study provides new evidence on the impact of information about relative performance on effort through the action of comparisons and peer effects. It would, however, be premature to draw general conclusions before testing whether our conclusions hold also in the context of a less cognitively demanding task. One could also provide other – perhaps more attractive – outside options, or even allow the quitters to leave the room as soon as they like. Another extension would consist of making the production period longer and investigate its impact on quitting decisions.

REFERENCES

- Aoyagi, M. (2007). Information Information in a Dynamic Tournament. Osaka University Institute of Social and Economic Research Discussion Paper No. 580.
- Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2005). Large Stakes and Big Mistakes. Federal Reserve of Boston Working Paper.
- Bandiera, O., Barankay, I., and Rasul, I. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics*, 120 (3), 917-962.
- Barber, B.M. and Odean, T. (2001). "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics*, 261-92.
- Baumeister, R.F. (1984). Choking under Pressure: Self-consciousness and Paradoxical Effects of Incentives on Skillful Performance. *Journal of Personality and Social Psychology*, 46(3), 610-20.
- Dohmen, T. (2008). Do Professionals Choke under Pressure? Forthcoming in *Journal of Economic Behavior and Organization*.
- Eckel, C. and Grossman, P. (2003). "Forecasting Risk Attitudes: An Experimental Study of Actual and Forecast Risk Attitudes of Women and Men." Mimeo.
- Ederer, F. (2004). Information and Motivation in Dynamic Tournaments. Mimeo, MIT.
- Ederer, F. and Fehr, E. (2007). Deception and Incentives: How Dishonesty Undermines Effort Provision? IZA Discussion Paper No.3200, Bonn.
- Ertac, S. (2005). Social Comparisons and Optimal Information Revelation - Theory and Experiments. Working Paper, UCLA.
- Falk, A. and Ichino, A. (2006). Clean Evidence on Peer Pressure. *Journal of Labor Economics*, 24 (1), 39-57.
- Fershtman, C. and Gneezy, U. (2005). Quitting. Discussion Paper No. 6-2005, Tel-Aviv University.
- Gershkov, A. and Perry, M. (2006). Tournaments with Midterms Reviews. Mimeo.
- Gibbs, M.J. (1991). An Economic Approach to *Process* in Pay and Performance Appraisals. Discussion Paper, Harvard Business School.
- Goltsman, M. and Mukherjee, A. (2006). Information Disclosure in Multi-Stage Tournaments. Mimeo.
- Gürtler, O. and Harbring, C. (2007). Information Tournaments under Commitment Problems: Theory and Experimental Evidence. IZA Discussion Paper No. 3111, Bonn.
- Guryan, J., Kroft, K., and Notowidigdo, M. (2007). Peer effects in the workplace; Evidence from random groupings in professional golf tournaments. NBER Working Paper No.13422.
- Haragushi, K.H. and Wadell, G.R. (2007). Interdependency in performance. IZA Discussion Paper No.2944, Bonn.
- Kandell, E. and Lazear, E.P. (1992). Peer Pressure and Partnerships. *Journal of Political Economy*, 100(4), p.801-17.
- Kluger, A.N. and Denisi, A. (1996). The Effects of Information Interventions on

Performance: A Historical Review, a Meta-analysis and a Preliminary Information Intervention Theory. *Psychological Bulletin*, 119, 254-284.

Kräkel, M. (2007). Emotions in Tournaments, forthcoming in *Journal of Economic Behavior and Organization*.

Lazear, E.P. (1989). Pay Inequality and Industrial Politics. *Journal of Political Economy*, (97), 561-580

Lazear E.P. (1992). The Job as a Concept, in: W. Burns (ed.), *Performance Measurement, Evaluation, and Incentives*. Boston: Harvard Business School Press, 183-215.

Lazear, E.P. and Rosen, S. (1981). Rank-order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5), 841-864.

Leonard, D. and Van Long, N. (2008). Is Emulation Good for You? The Ups and Downs of Rivalry. *CIRANO Working Paper*, 2008s-02.

Lizzeri, A., Meyer, M., and Persico, N. (2002). The Incentive Effects of Interim Performance Evaluations. CARESS Working Paper No.02-09.

Ludwig, S. and Ruchala, G. (2006). Knowing the Gap – Intermediate Information in Tournaments. Mimeo, University of Bonn.

Mas, A. and Moretti, E. (2006), Peers at Work. *American Economic Review*, forthcoming.

Milkovich G.T. and Newman, J.M. (1996). *Compensation*. Chicago: McGraw-Hill/Irwin

Müller, W. and Schotter, A. (2007). Workaholics and Drop Outs in Optimal Organizations. Mimeo, New York University.

Niederle, M. and Vesterlund, L. (2007). Do Women Shy away from Competition? Do Men Compete too Much? *Quarterly Journal of Economics*, August, 1067-1101.

O'Keefe, M., Viscusi, K.W., and Zeckhauser, R. (1984). Economic Contests: Comparative Reward Schemes. *Journal of Labor Economics*, 2, 27-56.

Schotter, A. and Weigelt, K. (1992). Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results. *Quarterly Journal of Economics*, 107, 511-539

Sprinkle, G.B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, 28(2,3), 287-318.



Yildirim, H. (2005). Contests with Multiple Rounds. *Games and Economic Behavior*, 51(1), 213-227.

Zeiliger, R. (2000). A Presentation of Regate, Internet Based Software or Experimental Economics. <http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt>, GATE.

Appendix 1. The screen in the Piece Rate-Continuous Feedback treatment (translated from French)

INCORRECT

Your score is the number of problems you have correctly solved. You have access to a total of 20 minutes.
Remaining time: **15 minutes, 43 seconds**

Your current score: 
Your co-participant's score: 

The numbers to add up are :

- 26
- 34
- 44
- 48

Enter your result:

Appendix 2 – Instructions for the Piece Rate - No Feedback treatment (original in French)

You are about to participate in an experiment in economics during which you can earn money. You will receive € for participating in this experiment. Your other earnings during this experiment are expressed in points, with the following conversion rate:

$$10 \text{ points} = 1 \text{ Euro}$$

At the end of the session, your earnings in points will be converted into Euros (rounded up) and added to the show-up fee. The total amount of the compensation you will receive will be paid to you in cash in private and in a separate room to preserve confidentiality.

Content of the task:

During this session, you are given the possibility to perform a task during a maximum of 20 minutes. This task consists of adding up four two-digit numbers that are randomly generated and that are displayed on your screen, like in the following example:

Example: 23
 45
 67
 12

Your task consists of adding up these numbers, that is $23+45+67+12$. The correct answer in this example is 147.

The use of paper, pencil, or calculator is forbidden. You must make these calculations in your head.

Once your calculation has been done, you enter your answer in the specified area and you submit your answer by clicking the "submit" button. You are immediately informed on whether your answer is correct or wrong. Whatever correct or wrong, a new series of numbers appears automatically. This can continue during 20 minutes.

During the experiment, you are free to read a book or a magazine that you have brought with you; you can also read the newspaper that has been put on your desk. In contrast, you must remain seated and silent until the end of the session.

Determination of payoffs:

You earn 1 point for each correct answer. You do not lose any point if you submit a wrong answer.

Information on your screen:

On your screen, you are permanently informed of the remaining time. You are also currently informed on your score (the current number of correct answers) and a cursor lengthens as soon as you submit a new correct answer.

After 20 minutes have elapsed, you will be informed of your score and the number of points you have earned.

Practice:

Before starting the experiment, we ask you to practice by adding up numbers during 3 minutes. This practice provides you with no additional point.

Then, after 20 minutes, we will ask you to answer a series of questions.

It is forbidden to communicate with the other subjects, or you would be excluded from the session. If you have any question regarding these instructions, please raise your hand. We will answer your question in private.

Appendix 3 – Instructions for the Tournament - Continuous Feedback treatment

You are about to participate in an experiment in economics during which you can earn money. You will receive € for participating in this experiment. Your other earnings during this experiment are expressed in points, with the following conversion rate:

$$10 \text{ points} = 1 \text{ Euro}$$

At the end of the session, your earnings in points will be converted into Euros (rounded up) and added to the show-up fee. The total amount of the compensation you will receive will be paid to you in cash in private and in a separate room to preserve confidentiality.

At the beginning of the session, all the subjects are randomly matched in pairs. Therefore, you are matched with another subject (your co-subject) throughout the session. The identity of this person will never be revealed to you.

Content of the task:

During this session, you are given the possibility to perform a task during a maximum of 20 minutes. This task consists of adding up four two-digit numbers that are randomly generated and that are displayed on your screen, like in the following example:

Example: 23
 45
 67
 12

Your task consists of adding up these numbers, that is $23+45+67+12$. The correct answer in this example is 147.

The use of paper, pencil, or calculator is forbidden. You must make these calculations in your head.

Once your calculation has been done, you enter your answer in the specified area and you submit your answer by clicking the "submit" button. You are immediately informed on whether your answer is correct or wrong. Whatever correct or wrong, a new series of numbers appears automatically. This can continue during 20 minutes.

During the experiment, you are free to read a book or a magazine that you have brought with you; you can also read the newspaper that has been put on your desk. In contrast, you must remain seated and silent until the end of the session.

Determination of payoffs:

You do not lose any point if you submit a wrong answer.

Your earnings depend on your score and on the score of your co-subject. After 20 minutes, your score is compared to the score of your co-subject. Three situations can occur.

- If your score is higher than your co-subject' score, you earn 170 points that are added to your show-up fee.
- If your score is lower than your co-subject' score, you earn 0 point. Thus, you only earn your show-up fee.
- If both scores are equal, we randomly determine who, between you and your co-subject, earns 170 points and who earns 0 point. These earnings are added to your show-up fee.

Information on your screen:

On your screen, you are permanently informed of the remaining time. You are also currently informed on your score (the current number of correct answers) and a cursor lengthens as soon as you submit a new correct answer.

Moreover, you are permanently informed of the score of your co-subject (the current number of correct answers of your co-subject) and a cursor lengthens as soon as your co-subject submits a new correct answer.

After 20 minutes have elapsed, you will be informed of your score, the score of your co-subject and the number of points you will earn.

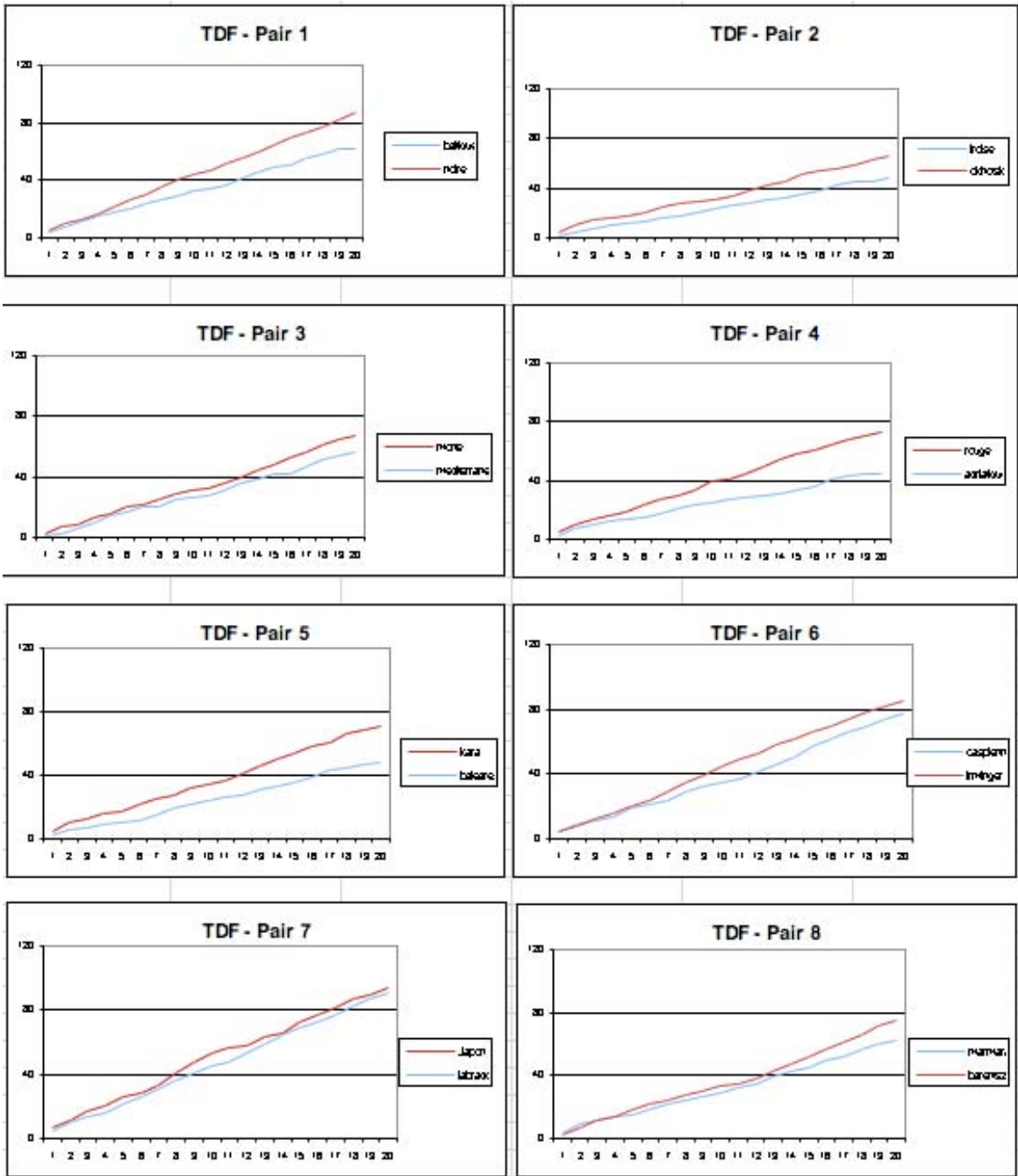
Practice:

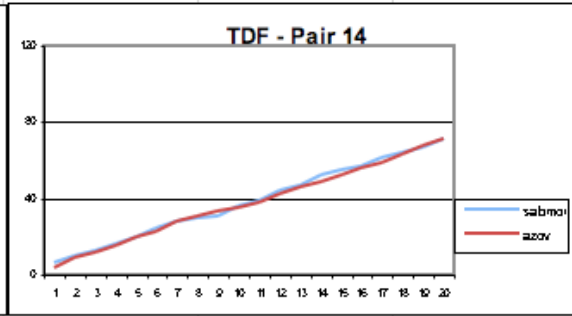
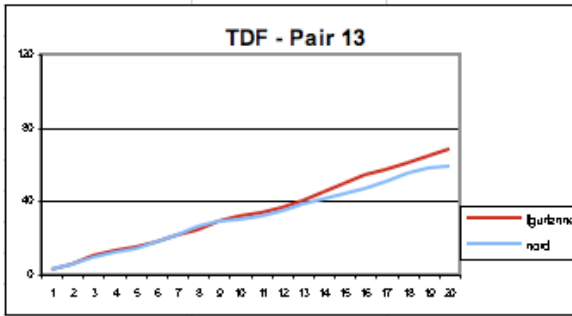
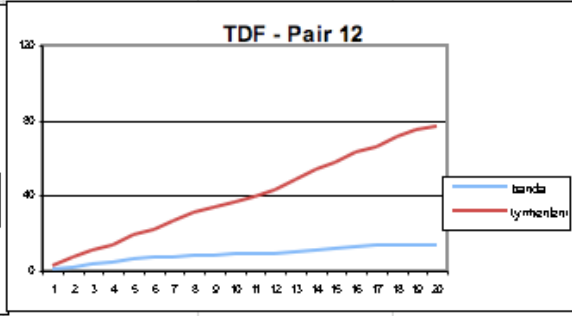
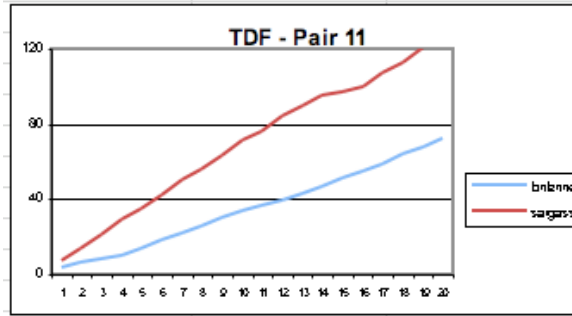
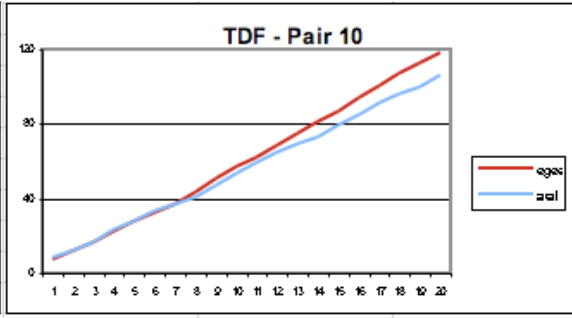
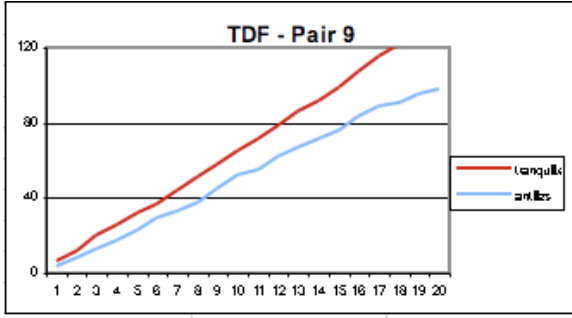
Before starting the experiment, we ask you to practice by adding up numbers during 3 minutes. This practice provides you with no additional point.

Then, after 20 minutes, we will ask you to answer a series of questions.

It is forbidden to communicate with the other subjects, or you would be excluded from the session. If you have any question regarding these instructions, please raise your hand. We will answer your question in private.

Appendix 4. Evolution of performance over time in the Tournament-Discrete Feedback treatment





Appendix 5. Evolution of performance over time in the Tournament-Continuous-Feedback treatment

