Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Competitive Incentives: Working Harder or Working Smarter?

Anat Bracha, Chaim Fershtman,

MANAGEMENT SCIENCE
Vol. 59, No. 4, April 2013, pp. 771–781
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1597

Competitive Incentives: Working Harder or Working Smarter?

Anat Bracha
Research Department, Federal Reserve Bank of Boston, Boston, Massachusetts 02210,
anat.bracha@bos.frb.org

Chaim Fershtman
The Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel 69978;
Centre for Economic Policy Research, London EC1V 3PZ, United Kingdom; and Tinbergen Institute,
3062 PA Rotterdam, The Netherlands, fersht@post.tau.ac.il

Almost all jobs require a combination of cognitive effort and labor effort. This paper focuses on the effect that competitive incentive schemes have on the chosen combination of these two types of efforts. We use an experimental approach to show that competitive incentives may induce agents to work harder but not necessarily smarter. This effect was stronger for women.

Key words: behavioral economics; individual decision making; lab experiment; competitive incentives; work effort
History: Received December 6, 2011; accepted May 14, 2012, by John List, behavioral economics. Published online in *Articles in Advance* September 21, 2012.

1. Introduction

Almost all jobs require a combination of cognitive effort and labor effort. Conducting research and development, running a company, building a house, teaching a class, or working on a factory floor requires both cognitive and labor efforts. The trade-off between these two types of effort exists whenever agents need to think about how to perform a task or to choose a method of solving a problem before they actually implement it. For example, consider the task of trying to find the highest value of a function: people may try to analyze the function (cognitive effort), they may try to check it for many parameter values (labor effort), or they may attempt a combination of the two methods.

The general intuition is that providing competitive incentives motivates individuals to exert more effort.[1] However, once we distinguish between cognitive and labor efforts, an additional question arises: What is the effect of competitive incentive on the *combination* of the two types of efforts? Clearly, there is not necessarily a trade-off between cognitive effort and labor effort because power incentives may induce agents to increase both of them.[2] But even in such cases, the question is whether agents choose to increase these types of efforts at the same proportion.

To answer this question, we examine players' effort allocation under two types of incentives: (i) a simple pay-for-performance incentive scheme (hereafter, PFP), in which agents are paid according to their own performance, and (ii) a tournament incentive scheme, in which pairs of participants compete for a prize. Our hypothesis is that competitive incentives will lead participants to exert relatively less cognitive effort but more labor effort relative to PFP.

Our hypothesis relates to the psychological literature identifying several mechanisms that result in "choking under pressure." This literature suggests that pressure in various forms, such as large stakes, performing in front of an audience, and competition, may lower performance in various types of tasks (see, e.g., Baumeister 1984, Baumeister and Showers 1986, Beilock et al. 2004). In economics, Ariely et al. (2009) demonstrated the choking-under-pressure effect and showed that excessively high rewards have a detrimental effect on performance.[3] Our hypothesis is based on the intuition that "it is difficult to think under pressure," and therefore competitive pressure

[1] For a survey, see Lazear (2000).

[2] See, for example, Kocher and Sutter (2006) for a study on the effect of time pressure that induces subjects to work both harder and smarter.

[3] However, choking under pressure does not necessarily occur in all circumstances. For instance, Beilock et al. (2004) showed that pressure adversely affects performance in solving novel, but not heavily practiced, math problems. Note that to correctly answer a practiced problem, one only needs to retrieve the solution from memory. Whereas solving novel math problems requires high working memory—cognitive effort—solving the practiced questions does not.

(as well as time pressure) has a negative effect on tasks that involve mainly cognitive effort.[4]

To test our claim, we designed a simple lab experiment with two computerized tasks: a "sequence" task where participants were asked to solve numeric sequences that require cognitive effort and a "filing" task that is a simple number-categorizing task mainly requiring manual dexterity. Participants in this study could engage in either task and were free to switch between the two during the entire duration of the experiment. To test our hypothesis, we measured time allocation between the two tasks and the success rate in solving sequences; we then compared these measures across conditions (i.e., PFP and tournament incentives).[5]

Because the overall score depends on the performance in the two tasks, we will use the following terminology. We say that players *work harder* whenever they reduce the time they spend on the cognitive task and use that time for the filing task. We say that players *work smarter* whenever they increase the proportion of time they spend on the cognitive task. Because subjects were given 10 minutes for the two tasks, we exclude, by construction, the possibility of working both harder and smarter. While our focus here is on time allocation, there might be other interpretations of working harder or smarter. Subjects may work harder in the filing task or in the sequence task by increasing the speed of filing or the speed of solving sequences. There are also alternative interpretations of what reflects working smarter, such as the success rate in solving sequences or choosing a better allocation of time between the two tasks.[6] We will report the effect of competitive incentives on these variables as well.

Our main results are that under competitive incentives, participants indeed devote less time to the sequence task and have a lower success rate than when they are provided with PFP incentives. We show that these results are not derived from risk aversion or from different optimal combination of tasks.

Our results, however, are gender sensitive. Under the PFP incentives, the performance of women is lower than that of men. Women attempted to solve fewer sequences and devoted more of the allotted time to the simpler task of categorizing numbers. This is despite the fact that in the PFP treatment, men and women had the same success rate in solving sequences. Analyzing the effect of competitive incentives by gender, we find that relative to the PFP treatment, competitive incentives induce both men and women to spend less time on the sequence task and more time on the routine filing task. However, the negative effect of the competitive incentives on the success rate is entirely a female effect.

This paper adds to the experimental investigation of tournament incentives relative to PFP incentives (see, e.g., Bull et al. 1987) and the recent literature on gender differences in response to competition (e.g., Gneezy et al. 2003; for a survey, see Croson and Gneezy 2009 and Bertrand 2011). Previous studies either used non-real-effort settings, where effort is a number chosen by a participant in the study, or a real-effort task that requires one type of effort. Hence, this paper adds to the literature by considering a new aspect of competitive incentives—the effect of competitive incentives on effort allocation. The paper also adds to the recent literature on the psychological foundation of incentives, which provides an important critical view of the traditional incentive theory (for a survey, see Fehr and Falk 2002). The main claim in this literature is that considering monetary incentives alone is too narrow, empirically questionable, and limits our understanding of incentives. Nonpecuniary motives such as reciprocity, the desire for social status, and fairness concerns are powerful drives of human motivation. Our results extend this literature by focusing on the combination of cognitive and labor effort.

Our paper is closely related to the literature on multitasking. In a seminal paper, Holmstrom and Milgrom (1991) considered optimal incentives in a multitask principal agent model.[7] Their focus was on a setup in which one task is observable and measured fairly accurately, whereas the second task cannot be measured or has a very noisy measurement. The agent needs to allocate his effort among the two tasks while only the first task is contractible.[8] The

[4] One possible interpretation of our intuition is that competitive pressure contributes to the agents' cognitive load, which makes it more costly to exert cognitive effort. Ariely et al. (2009) demonstrated that pressure (in the form of high rewards) may reduce performance in automated tasks because it induces players to think about what they are doing.

[5] The success rate is the percentage of sequences solved correctly over the number of sequences attempted.

[6] Note that even thinking about the optimal time allocation between the two tasks requires time and effort. Therefore players need to decide both how much time and effort to devote to deciding on time allocation and how much time and effort to devote to performing the tasks. One can clearly continue this argument for an additional layer of decisions, but these issues are beyond the scope of this paper.

[7] For a survey of this literature, see Prendergast (1999).

[8] As an example, Holmstrom and Milgrom (1991) described the controversy over the use of incentive pay for teachers based on their students' test scores. On the one hand, power incentives induce teachers to work harder; on the other hand, such incentives induce teachers to sacrifice important activities such as promoting curiosity and creative thinking in order to teach the narrowly defined basic skills that are tested in the standard exams.

main result of their analysis is that in such cases it would be better to pay fixed wages without any power incentive scheme.[9] Although our setting also has two tasks, our focus is different. In our setting the overall performance is determined by the combination of the two tasks, and only the overall performance is observable and contractible.[10] That is, the asymmetry between the activities is not with respect to observability but with respect to the nature of the activity itself, and for that reason, incentives do not have a distortionary effect in our setting.

Finally, it is important to note that not every lab result can be generalized to the outside world (see Levitt and List 2007). We follow the main points raised by Levitt and List to examine our ability to extrapolate our experimental finding beyond the lab to the real world. (i) Although in many lab settings morality issues or social norms may affect subjects' behavior, we do not think these are relevant in our lab experiment. It is not clear what the proper behavior is in the tasks they perform, and there are no moral costs that subjects tried to avoid.[11] (ii) The actions of subjects were scrutinized in our lab experiment. However, given the nature of our experiment, it is unlikely that subjects were aware of what we were trying to measure or of our research objectives. (iii) Context may indeed affect behavior, but we used a relatively neutral context. (iv) Selection is always a problem in a lab experiment. Our subjects were mostly Harvard students who are not representative of the entire population. One would therefore need to limit our results to the effect of tournaments on smart and educated subjects. (v) Clearly, as in most lab experiments, the stakes in our experiment were low. Low stakes are of a special concern when other preferences, such as prosocial preferences, are present. However, in our study we do not measure prosocial behavior, and there are no moral costs for one's actions. We therefore believe that the effect of small stakes is relatively limited.

2. Experimental Design

To capture the two different types of effort, we introduced two tasks. Subjects could engage in either solving sequences ("sequence" task) by finding a missing number in a sequence of four numbers or classifying a random number into an "odd" or "even" category ("filing" task) by pressing the appropriate button on the computer screen. The sequence task requires cognitive effort in the form of abstract thinking, whereas filing numbers mainly requires labor effort. Both tasks were available during the experiment, and engaging in each of the two tasks was done by simply clicking on the section of the screen with the desired task (see Figure 1). Subjects were given 10 minutes to work on the two tasks. We used a between-subject design with two treatments: pay-for-performance and head-to-head tournament.

PFP: Subjects were paid \$2 per *net* correctly solved sequences, 3 cents per *net* correctly filed numbers, and a 1 cent extra reward for the product obtained by multiplying the net sequences by the net filed numbers.[12] The net number of correctly solved sequences is the number of correctly solved sequences minus half the number of incorrectly solved sequences. Penalizing incorrectly solved sequences was designed to prevent guessing. The net filed numbers equals the correctly filed numbers minus the incorrectly filed numbers. Penalizing incorrectly filed numbers was designed to prevent random clicking. The extra reward introduces a complementary term, as a greater number of net correctly solved sequences (filed numbers) increases the marginal return of successful filing (sequence).

Tournament: In this treatment, subjects were randomly paired using a randomly generated subject identification (ID) number. The pairs were announced before the beginning of the task and by subject ID, such that the identity of one's opponent was not revealed. The winner was determined according to the accumulated number of points for each of the opponents in a pair. The point schedule was exactly as under the PFP compensation scheme. The winner's prize was \$60, and the loser received the minimum guarantee of \$10, such that the expected earning was \$35, similar to the average earning under PFP. At the end of the study, after completing the time devoted to the task, the accumulated number of points for each participant was announced (by the randomly generated subject ID), and the earnings were determined and announced.

2.1. Procedure

The sessions were conducted at the Harvard Decision Science Laboratory at the Harvard Kennedy School. A total of 134 Harvard students participated in the study, 74 in the PFP treatment and 60 in the tournament treatment. There were 48 females and

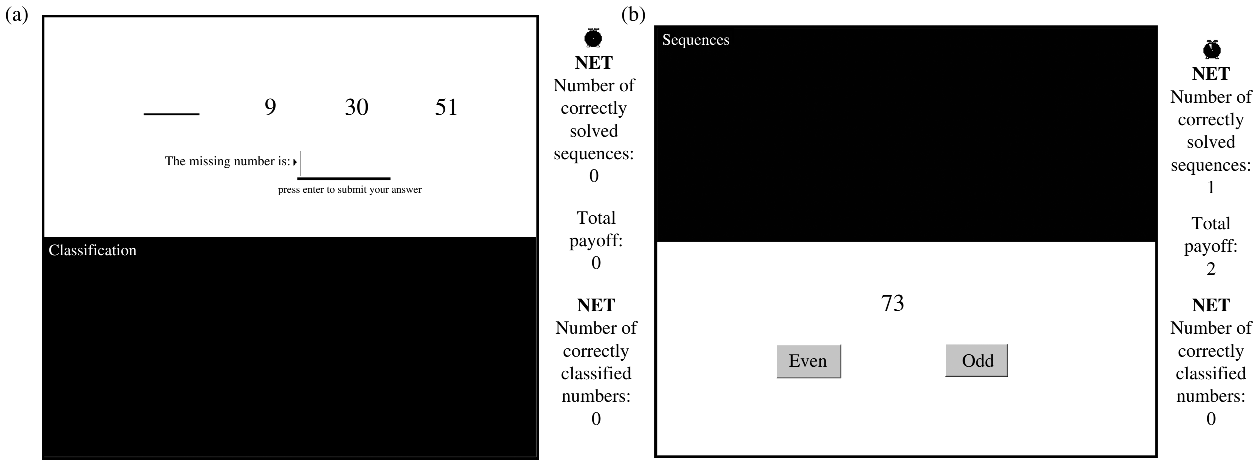
[9] For an experimental study of this multitask principal agent model, see Fehr and Schmidt (2004) and Al-Ubaydli et al. (2011).

[10] When the two tasks are observable and contractible, it is possible, in principle, to provide incentives for each task separately. However, the focus of our research is on jobs that can be performed by different combinations of cognitive and labor efforts, and this combination is not contractible. In our experiment the two tasks are observable as the chosen mix of such tasks is the focus of our paper.

[11] There might, however, be identity cost associated with our two tasks. That is, subjects may choose to spend more time on the cognitive task simply because they view themselves as smart individuals that can face the cognitive challenge. These considerations may exist in the lab as well as in the outside world.

[12] This compensation is different from a piece rate because there is a multiplicative term in the incentives.

Figure 1 **Sequence Task (a) and Filing Task (b)**



26 males in the PFP treatment and 35 females and 25 males in the tournament treatment. In each session, participants sat at individual computer stations and read the instructions on their individual screens. Once all subjects finished reading the instructions, they were given a code to proceed such that all subjects started working on the task at the same time. In the tournament treatment, once all subjects finished reading the instructions, the experimenter announced the pairs by subject ID before giving the code to proceed.

3. Competitive Incentives

Economic intuition suggests that competitive incentives may induce individuals to exert more effort.[13] However, individuals may also react to the pressure created by the competitive incentives. Competitive pressure may have a different effect on cognitive and labor effort. Our main hypothesis is that competitive pressure affects mainly cognitive effort, making it more difficult or more costly to perform cognitive tasks.

Hypothesis 1. (i) *Competitive incentives induce individuals to reduce their cognitive effort while increasing their labor effort*. (ii) *Competitive incentives reduce the effectiveness of cognitive effort*, *which results in a lower success rate*.

In our experiment individuals were given 10 minutes to engage in the two tasks. This time limit may have created an additional type of pressure—time pressure. We expect that the time pressure, if it exists, will be stronger at the end of the given time frame and that time pressure will be stronger in the tournament treatment because an additional point may be crucial for winning the tournament.

Hypothesis 2. *Time pressure will be effective in the tournament treatment*, *resulting in lower success rates at the end of the tournament relative to those at the beginning of the tournament*.[14]

To test these hypotheses, we examined the time allocation between the two tasks and the success rate in solving sequences. The comparison is provided in Table 1.[15]

3.1. The Effect of Competitive Incentives

The effect of tournament incentives on time allocation can be seen in Figures 2(a) and 2(b). It is best seen in Figure 2(b), which compares the minute-by-minute percentage of time devoted to solving sequences in the PFP and the tournament treatments. Figure 2(b)

[13] See, e.g., Lazear and Rosen (1981) for a theoretical argument. Empirically, Ehrenberg and Bognanno (1990) demonstrated the positive effect of tournament incentives on effort in golf tournaments, and Kremer et al. (2009) found a positive effect of scholarship competition for girls in Kenya on their school achievements. Experimentally, results of studies such as Gneezy et al. (2003) and Niederle and Vesterlund (2007), although aimed at gender differences, show that competitive incentives increase performance among men relative to performance with piece rate, and even among women, performance is at least as good under tournament as it is under piece rate.

[14] Our hypothesis is only with respect to success rate and not to the allocation of time between the two tasks because the allocation of time at the end is affected by the decisions at the beginning of the 10 minutes.

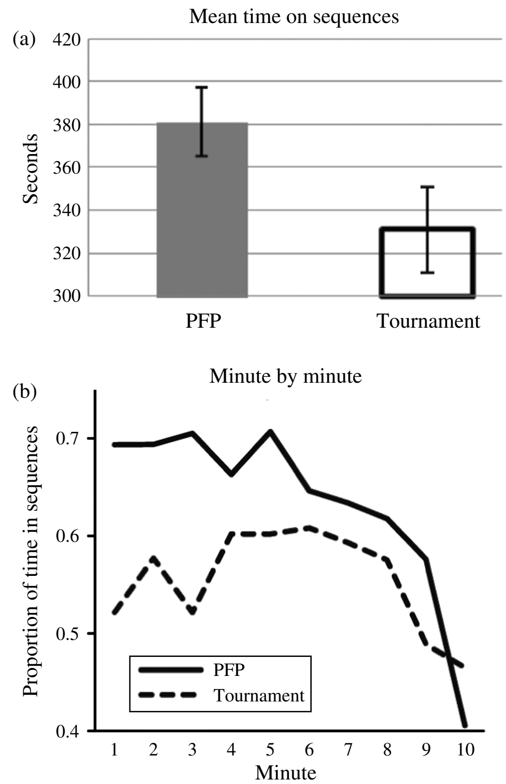
[15] In calculating the various averages, we first calculated the particular measure (such as success rate) for each individual and then averaged across individuals. Under PFP incentives, the average number of sequences attempted was 11 with a standard deviation of 6.69; under tournament, the average number of sequences attempted was 9.6 with a standard deviation of 6.78. There were two outliers, one in each condition, who attempted more than 30 sequences in 10 minutes (32 sequences under PFP and 33 sequences under tournament); 30 attempts is more than two standard deviations from the mean. Therefore, in our analysis and in the statistics presented above, we exclude these two outliers.

Table 1 Comparing the PFP and Tournament Treatments

	PFP	Tournament	*p*-value
Total performance	33.95	32.99	n.s.
Attempted sequences	10.71	9.2	0.083
Correctly solved sequences	8.54	7.29	n.s.
Success rate	0.79	0.72	0.047
Time on sequences	381.06	330.88	0.025
Time per attempted sequence	48.67	50.79	n.s.
Time per correctly solved sequence	63.42	70.74	n.s.
Net filing rate	1.02	1.06	n.s.

Notes. Success rate, time per attempted sequence, and time per correctly solved sequence are based on individuals who actually solved a sequence. The *p*-values are based on one-sided *t*-tests.

Figure 2 Allocation of Time



Notes. $N = 73$ under PFP, $N = 59$ under tournament. The bars in (a) represent the standard error of the mean.

strikingly illustrates that the effect is neither due to a single episode nor due to a particular stage of the task. The effect of tournament incentives on time allocation stems from different time allocations made throughout the entire 10 minutes of the study.[16]

To complete the investigation of Hypothesis 1, we calculated the success rate in solving sequences for each participant and then averaged these results across all individuals (see Figure 3(a)). We find that the success rate is, indeed, lower under competitive incentives. Labor effort can be captured by examining the average net filing rate, which is the average speed of net filing across participants (see Figure 3(b)).[17]

Observation 1 (Tournament and Cognitive Effort). (i) Under competitive incentives participants devoted less time to the cognitive task than under the PFP incentives (330 seconds under competition versus 381 under PFP; $p = 0.025$). (ii) The success rate in solving sequences is lower under tournament incentives—78.6% under PFP and only 72% under tournament incentives ($p = 0.047$).

Clearly, the two parts of Observation 1 may be interdependent. If participants are aware of the fact that under the tournament incentive scheme they have a lower success rate, their rational reaction would be to reduce the time spent on sequence solving. The effect of competitive incentives on subjects' success rate may be the outcome of two different effects. Subjects may reduce the cognitive effort in the sequence task that results in a lower success rate. However, it is also possible that competitive pressure affects the players' ability. The conservative economic approach assumes that ability is exogenously given and not affected by incentives. Nevertheless, it is conceivable that ability would be lower as a result of competitive pressure. Under such an approach, the effect of incentive schemes on ability should be taken into account whenever incentives are designed.[18]

One possible explanation for Observation 1(i) is that it is indeed optimal for players to reduce the number of sequences they attempt to solve under tournament incentives. To clarify this point, we calculated the optimal number of sequences for participants in the tournament condition using our data on subjects' net filing rate, speed of solving a sequence, and the success rate.[19] We first calculated the optimal number of sequences when the objective is to win

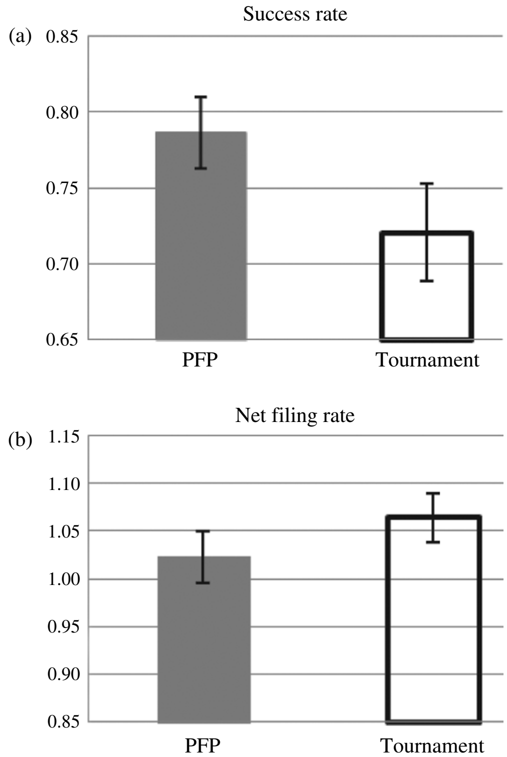
[16] To test whether the difference in time allocation between the two treatments decreases over time, we ran an ordinary least squares (OLS) regression of the percentage of time spent on sequences on minute (1–10), treatment (=1 for tournament), and their interaction. We find a significant negative effect of tournament ($p < 0.01$) that declines with time (the interaction term is positive and significant, $p < 0.05$).

[17] The net filing rate is defined as the amount of the numbers filed correctly minus the amount of numbers filed incorrectly divided by the time spent on the filing task.

[18] The standard setting in the economic literature is to view abilities as given. This basic assumption simplifies contract theory because it allows us to focus only on the effect of incentives on players' decisions. A more radical view may argue that incentive schemes may affect the players' abilities as well.

[19] We have this information for all the participants who attempted to solve at least one sequence and who filed at least one number.

Figure 3 **Sequence Success Rate**



Notes. In (a), $N = 73$ under PFP; $N = 57$ under tournament. In (b), $N = 69$ under PFP; $N = 56$ under tournament. Bars represent the standard error of the mean. There are fewer observations when looking at the net filing rate compared to success rate because some subjects did only sequences whereas others did only filing.

the tournament and then compared it to the calculated optimal number of sequences to maximize payoff under the PFP incentives schedule.

When calculating the optimal number of sequences for winning the tournament, we calculated for each player and for every possible number of sequences she could solve in the given time frame her distribution of points. We then let each player compete against the distribution of points that players actually achieved in the tournament treatment (excluding the considered player) and then calculated the player's probability of winning. The player's optimal number of sequences is the number that maximizes her probability of winning. We did this for all players who attempted to both solve sequences and file numbers, and we found that the average optimal number of sequences (that players should attempt to solve) is 7.82. We then calculated the optimal number of sequences for maximizing payoff under the PFP incentives, that is, the number of sequences that players need to choose if they wish to maximize the expected number of points. The average optimal number of sequences under the PFP incentives is 7.72 sequences. That is, the average optimal number of sequences under tournament is slightly (but not significantly) higher than the average optimal number of sequences under PFP incentives. Therefore, Observation 1(i) cannot be explained by players' optimal response to the different incentive schemes.

3.2. Time Pressure

The 10-minute time limit in our experiment might give rise to time pressure—in particular, just before the end of the session. We therefore divide the 10-minute experiment into two parts: the first 7 minutes and the last 3 minutes, and we compare the participants' success rate at the beginning and at the end of the treatment.[20]

In the tournament treatment, the success rate was 77.5% during the first seven minutes and only 57.9% during the last three minutes. This decline is highly significant ($p < 0.01$). In the PFP treatment, the success rate in the first seven minutes was 76.4% and 77.2% in the last three minutes. Of course, it is possible that in the tournament treatment some low-ability individuals chose to solve sequences only during the last three minutes. To exclude this possibility, we compared the success rate only for those individuals who solved sequences during both the first seven minutes and the last three minutes. We find the same pattern: there was no effect under PFP (76.1% success rate during the first seven minutes, and 77% success rate during the last three minutes). On the other hand, there was a highly significant decline under tournament incentives (76.3% success rate during the first seven minutes compared with 57.5% success rate during the last three minutes; $p < 0.01$).

Observation 2 (Competitive Time Pressure). During the last three minutes of the tournament, the participants' success rate was significantly lower than in the first seven minutes. However, time pressure in and of itself had no such effect, as in the PFP treatment there was no reduction in participants' success rate during the last three minutes.

3.3. Tournament and Risk Aversion

Solving sequences is a risky task, as players are uncertain about whether they will be able to solve a sequence correctly and about how long it will take them to solve a sequence. In the PFP treatment, risk aversion may be an important factor in determining the time allocation between the two tasks. The question is whether the riskiness of the cognitive task drives the result reported in Observation 1.

In evaluating the overall risk that subjects faced in our experiment, one needs to separate the risk derived from the task they choose and the risk that

[20] This 7:3 division is arbitrary.

is derived from the incentives they face. In the PFP treatment, subjects were paid according to their own performance; hence, when they engaged in the filing task, they bore no risk. However, in the tournament treatment, subjects faced the risk of losing the tournament, regardless of the task they chose. Under such an incentive scheme, even the filing task yields uncertain outcome. As we will illustrate below, it is not clear that the "risky" task is riskier than the "safe" task when players are facing tournament incentives.

Suppose a player has two choices. Choice A yields 22 for sure, and choice B gives 35 with probability 50% and 10 with probability 50%. Under the PFP incentives, where the numbers are dollar amounts, choice B is risky whereas choice A is not. Whenever players are (sufficiently) risk averse, we would expect them to prefer choice A over choice B. Now suppose that players are divided into pairs and face tournament incentives; the numbers now represent the number of points one can achieve. The player who earns the highest number of points wins the tournament and gets a prize; in case of a tie, the winner is determined by a flip of a fair coin. In such a competitive environment, although choice A seems safe (22 points for sure), it is in fact as risky as choice B. Choosing either one—choice A or choice B—the individual bears the same exact risk. The probability of winning is 50% for each choice of action regardless of the choice of the player's opponent. For a similar reason, the filing task in our case is not risky under PFP incentives but yields uncertain returns under competitive incentives.

We ran a simple choice experiment to illustrate the above argument. The experiment was conducted at the Harvard Decision Science Laboratory with 50 participants—20 males and 30 females—in three sessions. We use a within-subject design, where participants were given two choices: choice A guaranteed a fixed amount of 22, whereas choice B was a lottery of 50/50 chance to win 35 or 10. In one task the amounts were marked in dollars and represented their final payoffs (equivalent to our PFP treatment). In the second task players were involved in a head-to-head tournament, and the amounts were marked in points. The player that got the highest amount of points was the winner and got the prize of $40 (in case of a tie, we flipped a fair coin). The order of the tasks was different across sessions, and in all sessions, one of the choice tasks was randomly picked for payment at the end of the study.

Notice that in both tasks, the choice is between a sure amount and a lottery, where the amounts and the lottery chances were identical across the two tasks. The only difference is the environment—in one task one is paid according to her choice, whereas in the other the payment depends on a competition result. This is similar to the two tasks in the original study—the filing task seems "safe" while the sequence task seems "risky." If competition leads individuals to choose what seems safe, we would expect a higher percentage of individuals to choose the sure amount (choice A) under tournament incentives than under the PFP incentives. The findings of the experiment were just the opposite. Participants chose the lottery more often under tournament incentives. Under tournament, 68% of the subjects chose the risky option; under PFP incentives, only 24% chose the risky option ($p < 0.01$). This is true for both males (35% versus 75%; $p < 0.01$) and females (17% versus 63%; $p < 0.01$).

Table 2 **Winners and Losers in the Tournament Treatment**

	Winners	Losers	*p*-value
Total performance	44.38	21.21	0.00
Attempted sequences	11.13	7.20	0.001
Correctly solved sequences	10	4.5	0.00
Success rate	0.85	0.59	0.00
Time on sequences	347.83	313.35	n.s.
Time per attempted sequence	47.56	54.14	n.s.
Time per correctly solved sequence	56.18	86.97	0.052
Net filing rate	1.10	1.02	n.s.

Notes. For winners, $N = 30$; for losers, $N = 29$. Success rate, time per attempted sequence, and time per correctly solved sequence are based on individuals who actually solved a sequence. The *p*-values are based on two-sided *t*-tests.

3.4. Winners and Losers

Competitive incentives do not necessarily have a uniform effect on individuals: some people perform better under competition, while others may get discouraged and perform worse due to competitive pressure. We therefore examined the effect of competitive incentive by subgroups. We first distinguished between winners and losers in the tournament treatment and then compared high and low performers in the tournament treatment to the appropriate comparison group in the PFP treatment.

Table 2 presents the performance of the winners and the losers separately. We find that winners and losers spend statistically the same amount of time on solving sequences—the winners spend 348 seconds, on average, whereas losers spend 313 seconds, on average, but this difference is not significant ($p = 0.39$). Nevertheless, the winners' average score is 44.38 points, they solve on average 11.13 sequences, and their success rate is 85%; the losers' average score is 21.21 points, they solve only 7.20 sequences, and their success rate was 59%.

Observation 3. There was no statistical difference between the tournament winners and losers with respect to the time they spent on solving sequences.

The winners in the tournament treatment are the participants with the higher success rate.

Next we split the PFP and the tournament participants into two groups each—above- and below-median performers, based on overall performance in that specific condition.[21] We found that the success rate of the above-median performers across conditions is similar (85.64% under tournament versus 84.96% under PFP). However, the success rate of the below-median performers is affected by the competitive environment: the success rate of the below-median performers in the tournament is 57.93%, and it is significantly lower than the 72.45% success rate of the below-median PFP performers ($p < 0.05$). As for time allocation, both the above- and below-median performers in the tournament seemed to spend less time solving sequences compared with above- and below-median PFP performers, respectively. Below-median performers in the PFP condition spent 366 seconds on sequences, while below-median performers in the tournament spent only 315 seconds on sequences. Similarly, above-median performers in the PFP condition spent 397 seconds on sequences, while above-median performers spent only 347 seconds on sequences under tournament. However, the decline is significant only for the above-median performers ($p = 0.064$).

Observation 4. The decline in the time spent on sequences under competitive incentives is evident across performance levels; however, it is significant only among the high performers. The decline in the success rate, in contrast, characterizes only the low performers.

4. The Gender Effect

Recent studies have indicated that men and women respond differently to competitive incentives.[22] Much to our surprise, not only did we find gender differences in response to competitive incentives, we also found gender differences in the benchmark PFP treatment.

Observation 5 (PFP: Gender Effect). (i) Women devoted, on average, 360 seconds to solving sequences while men devoted 419 seconds to solving sequences ($p = 0.055$).[23]

(ii) On average, women attempted to solve 8.9 sequences while, on average, men attempted to solve 14.1 sequences ($p < 0.01$).

(iii) Women spent on average 50.79 seconds per attempted sequence while men spent on average 44.58 seconds per attempted sequence (this difference is not statistically different).

(iv) Men and women had similar average success rates: 76.2% for men and 79.8% for women.[24]

Note that in the PFP treatment, women and men had a similar success rate; in fact, although not statistically significant, the women's average success rate was slightly higher than the men's. Thus, the women's choice to devote less time to solving sequences cannot be the outcome of a lower success rate.

It is possible that the observed time allocation choice is the outcome of gender differences in risk aversion and self-confidence (see, e.g., Eckel and Grossman 2008a; for a review, see Eckel and Grossman 2008b, as well as Croson and Gneezy 2009 and Bertrand 2011). As discussed, under PFP incentives, solving sequences is riskier than filing numbers. Therefore a higher degree of risk aversion among women is a possible explanation for why women devote less time to sequence solving compared with men. Beyond risk aversion, lower self-confidence, which can affect one's own perception of success rate and time needed to solve sequences, may also be a factor explaining women's choice of spending less time on sequences.

The overall performance was not different in the tournament and PFP treatments: the average number of points for women under tournament was 29.12 compared with $30.98 under PFP. The average number of points for men under tournament was 38.64 compared with $39.64 under PFP. The differences for both women and men are insignificant.

[21] Since we had a random matching of pairs in the tournament treatment, the losers and winners are not necessarily all of low- or high-ability, respectively. This is because it is possible that two strong or two weak participants were competing against each other. For that reason we compare low (high) performers in the PFP condition to the low (high) performers in the tournament condition.

[22] See Gneezy et al. (2003) for gender differences in response to competition among college students and Gneezy and Rustichini (2004) for gender differences in response to competition among children. See Niederle and Vesterlund (2007), Sutter and Rützler (2010), and Datta Gupta et al. (2011) for gender differences in selecting into competitive environment. Gneezy et al. (2009) suggested that these differences may be due to nurture rather than nature; see also Booth and Nolen (2012). For recent reviews, see Croson and Gneezy (2009) and Bertrand (2011). In more recent papers, Dreber et al. (2011) found a significant gender difference in competitiveness among Swedish children, whereas Cárdenas et al. (2012) found an insignificant gap among children in Colombia.

[23] The reader may wonder whether the observed effect of tournament incentives reported above is an artifact of having a different gender mix across treatments. However, because there were relatively more women participating in the PFP treatment than in the tournament treatment, we would expect the opposite effect.

[24] There was no gender difference in the speed of the filing task—women's average net filing rate was 1.00, and for men, it was 1.04.

4.1. The Effect of Competitive Incentives on Women

The women's success rate in solving sequences declined from 79.8% under PFP to 67.1% under tournament ($p < 0.01$), a sharp and strong decline of over 15%. This decline was evident in the last three minutes of the experiment—a decline from 77.1% under PFP to 49.7% under tournament ($p < 0.01$)—but was not evident during the first seven minutes. Under the PFP incentives, women spent an average of 360 seconds on solving sequences, while under tournament incentives they spent, on average, only 308 seconds ($p = 0.066$).[25] See Figure 4(a) for the minute-by-minute time allocation in the PFP and tournament treatments.

4.2. The Effect of Competitive Incentives on Men

In the tournament treatment, men reduced the amount of time they devoted to solving sequences from 419 seconds to 363 seconds ($p = 0.059$; for the minute-by-minute time allocation to sequences, see Figure 4(b)). The average number of sequences men attempted to solve decreased from 14.16 under PFP to 10.87 under tournament ($p = 0.045$). However, the tournament incentives did not affect men's success rate, which was 76.2% in the PFP treatment and 78.6% in the tournament treatment.

Observation 6 (Gender and Tournament). (i) Both men and women reduced the time they spent on sequence solving when facing tournament incentives. (ii) Tournament incentives only affected the success rate of women. This effect is mainly due to pressure at the end of the tournament.

These results are also reflected in a regression analysis that controls for gender and age: we used an OLS regression of the success rate and the time devoted to solving sequences on a treatment dummy variable (that takes a value of 1 for a tournament treatment), gender (that takes a value of 1 for females), and age. The results are presented in Table 3.

As Table 3 shows, we find that under competitive incentives, the success rate is lower by 7.5 percentage points, which is approximately 9.5% of the average success rate under PFP ($p = 0.06$). Under tournament incentives, the time allocated to solving sequences is lower by 52.38 seconds, which is about 13.7% of the average time devoted to sequences under PFP ($p = 0.038$). Adding an interaction term to examine whether competition has a differential gender effect confirms that the decline in the success rate under tournament is solely a female effect ($p = 0.06$), whereas the decline in the time allocated to solving sequences is similar for both men and women (the interaction

[25] This led to a significantly higher net filing among women (339 under tournament versus 258 under PFP; $p = 0.024$).

Figure 4 **Allocation of Time (Minute by Minute)**

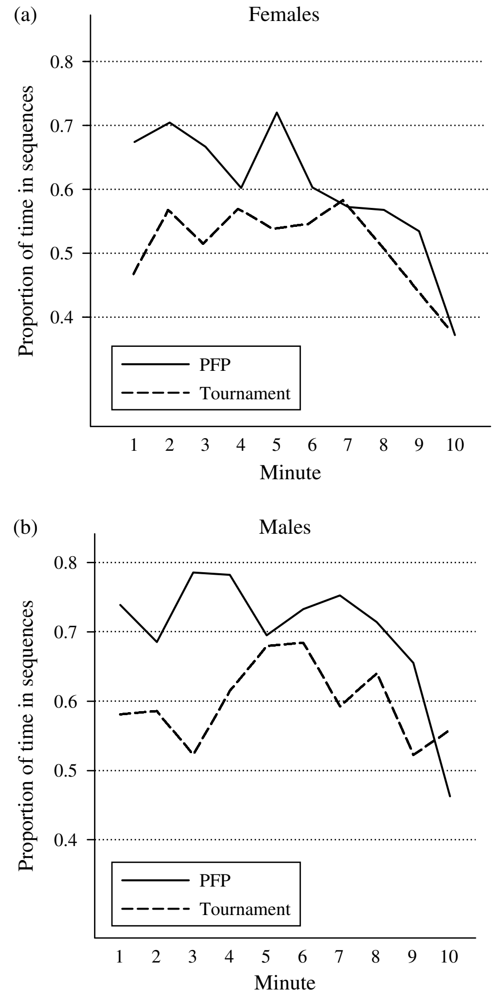


Table 3 **OLS Regressions**

	Success	Time allocated to sequences	Success	Time allocated to sequences
Treatment	**−0.075**	**−52.381**	0.017	−59.322
(=1 for tournament)	**(0.06)**	**(0.03)**	(0.78)	(0.14)
Gender	−0.030	−53.030	0.038	−58.177
(=1 for females)	(0.44)	(0.04)	(0.48)	(0.10)
Treatment × *Gender*			**−0.150**	**11.157**
			(0.06)	**(0.83)**
Age	−0.018	−7.209	−0.017	−7.268
	(0.23)	(0.06)	(0.23)	(0.06)
Constant	1.179	564.671	1.130	569.255
	(0.00)	(0.00)	(0.00)	(0.00)
N	130	132	130	132
R^2	0.037	0.089	0.064	0.089

Notes. The *p*-values are in parentheses; the number of observations is lower when analyzing the success rate compared with time allocation. This is due to participants who did not try to solve a single sequence.

term is not significant). Furthermore, testing whether the overall treatment effect is significant for females, we find a significant treatment effect on success rate (Wald test, $p < 0.01$), but not on time allocation.

Interestingly, in our settings women won the tournament at a similar proportion as did men, in contrast to previous findings (see Table A.1 in the appendix; see Gneezy et al. 2003, Gneezy and Rustichini 2004). Specifically, 16 out of the 35 women who participated in the tournament treatment won, whereas 14 out of 24 men who participated in the tournament treatment won (Fisher exact test, $p = 0.43$). Furthermore, by randomly matching participants in the tournament condition and repeating this test, we find that out of 100 random matchings there were only 8 instances with significant differences in the winning proportions across genders.[26] The result is that we reject the hypothesis that the average z-statistics across all 100 random matches is equal to or greater than 1.96. Therefore, this lack of difference in the winning proportions across gender is not an artifact of the actual matching we used in the study.[27]

5. Concluding Comment

Modern organizations typically provide workers, managers, students, or researchers with strong and competitive incentives to induce them to exert more effort. However, there are several studies showing that this intuitive effect does not always hold. For example, Gneezy and Rustichini (2000) and Frey and Jegen (2001) demonstrated the crowding-out effect, where strong explicit incentives may crowd out social motivation and may result in less effort (for a recent survey, see Gneezy et al. 2011). The main result of this paper focuses on yet another shortcoming of strong competitive incentives—they may induce agents to work harder but not necessarily smarter.

Acknowledgments

The authors thank department editor John List, the associate editor, and three anonymous referees for many valuable comments and suggestions. The authors thank Lynn Conell-Price for her excellent research assistance. The views expressed in this paper are solely those of the authors and do not reflect those of the Federal Reserve System or the Federal Reserve Bank of Boston. This work was supported by the Israel Science Foundation [Grant 0610114192].

[26] There were, in total, 100 random matchings of the entire group of participants.

[27] In generating the random matching, we tried (1) using all participants, including those who attempted over 30 sequences, and (2) excluding those who attempted over 30 sequences. In the latter case, we simply dropped the person left with no competitor, and we found similar results.

Appendix

Table A.1. Winners and Losers in the Tournament Treatment, by Gender

	Winners	Losers	p-value
Total performance	44.38	21.21	0.00
Women	39.75	20.17	0.00
Men	49.67	23.18	0.00
Success rate	0.85	0.59	0.00
Women	0.80	0.56	0.00
Men	0.89	0.64	0.00
Time on sequences	347.83	313.35	0.39
Women	302.97	313.37	0.85
Men	399.11	313.30	0.14
Net filing rate	1.10	1.02	0.13
Women	1.12	1.00	0.057
Men	1.07	1.06	0.88

Notes. For winners, $N = 30$ (14 men, 16 women). For losers, $N = 29$ (10 men, 19 women).

References

Al-Ubaydli O, Andersen S, Gneezy U, List JA (2011) Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. Working paper, George Mason University, Fairfax, VA.

Ariely D, Gneezy U, Loewenstein G, Mazar N (2009) Large stakes and big mistakes. *Rev. Econom. Stud.* 76(2):451–469.

Baumeister RF (1984) Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *J. Personality Soc. Psych.* 46(3):610–620.

Baumeister RF, Showers CJ (1986) A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *Eur. J. Soc. Psych.* 16(4):361–383.

Beilock SL, Kulp CA, Holt LE, Carr TH (2004) More on the fragility of performance: Choking under pressure in mathematical problem solving. *J. Experiment. Psych.: General* 133(4):584–600.

Bertrand M (2011) New perspectives on gender. Ashenfelter O, Card D, eds. *Handbook of Labor Economics*, Vol. 4B (North-Holland, Amsterdam), 1543–1590.

Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *J. Econom. Behav. Organ.* 81(2):542–555.

Bull C, Schotter A, Weigelt K (1987) Tournaments and piece rates: An experimental study. *J. Political Econom.* 95(1):1–33.

Cárdenas J-C, Dreber A, Von Essen E, Ranehill E (2012) Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *J. Econom. Behav. Organ.* 83(1):11–23.

Croson R, Gneezy U (2009) Gender differences in preferences. *J. Econom. Literature* 47(2):448–474.

Datta Gupta N, Poulsen A, Villeval MC (2011) Gender matching and competitiveness: Experimental evidence. *Econom. Inquiry*, ePub ahead of print April 26, http://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.2011.00378.x/abstract.

Dreber A, von Essen E, Ranehill E (2011) Outrunning the gender gap—Boys and girls compete equally. *Experiment. Econom.* 14(4):567–582.

Eckel CC, Grossman PJ (2008a) Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *J. Econom. Behav. Organ.* 68(1):1–17.

Eckel CC, Grossman PJ (2008b) Men, women and risk aversion: Experimental evidence. Plott CR, Smith VL, eds. *Handbook of Experimental Economics Results* (Elsevier, Amsterdam), 1061–1073.

Ehrenberg RG, Bognanno ML (1990) Do tournaments have incentive effects? *J. Political Econom.* 98(6):1307–1324.

Frey BS, Jegen R (2001) Motivation crowding theory. *J. Econom. Surveys* 15(5):589–611.

Fehr E, Falk A (2002) Psychological foundation of incentives. *Eur. Econom. Rev.* 46(4–5):687–724.

Fehr E, Schmidt KM (2004) Fairness and incentives in a multi-task principal-agent model. *Scand. J. Econom.* 106(3):453–474.

Gneezy U, Rustichini A (2000) Pay enough or don't pay at all. *Quart. J. Econom.* 115(3):791–810.

Gneezy U, Rustichini A (2004) Gender and competition at a young age. *Amer. Econom. Rev. Papers Proc.* 94(2):377–381.

Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77(5):1637–1664.

Gneezy U, Meier S, Rey-Biel P (2011) When and why incentives (don't) work to modify behavior. *J. Econom. Perspect.* 25(4): 191–210.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quart. J. Econom.* 118(3):1049–1074.

Holmstrom B, Milgrom P (1991) Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *J. Law, Econom. Organ.* 7(Special issue):24–52.

Kocher MG, Sutter M (2006) Time is money—Time pressure, incentives, and the quality of decision making. *J. Econom. Behav. Organ.* 61(3):375–392.

Kremer M, Miguel E, Thornton R (2009) Incentives to learn. *Rev. Econom. Statist.* 91(3):437–456.

Lazear EP (2000) The future of personnel economics. *Econom. J.* 110(467):611–639.

Lazear EP, Rosen S (1981) Rank-order tournaments as optimum labor contracts. *J. Political Econom.* 89(5):841–864.

Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect.* 21(2):153–174.

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

Prendergast C (1999) The provision of incentives in firms. *J. Econom. Literature* 37(1):7–63.

Sutter M, Rützler D (2010) Gender differences in competition emerge early in life. IZA Discussion Paper Number 5015, IZA, Bonn, Germany. http://ftp.iza.org/dp5015.pdf.