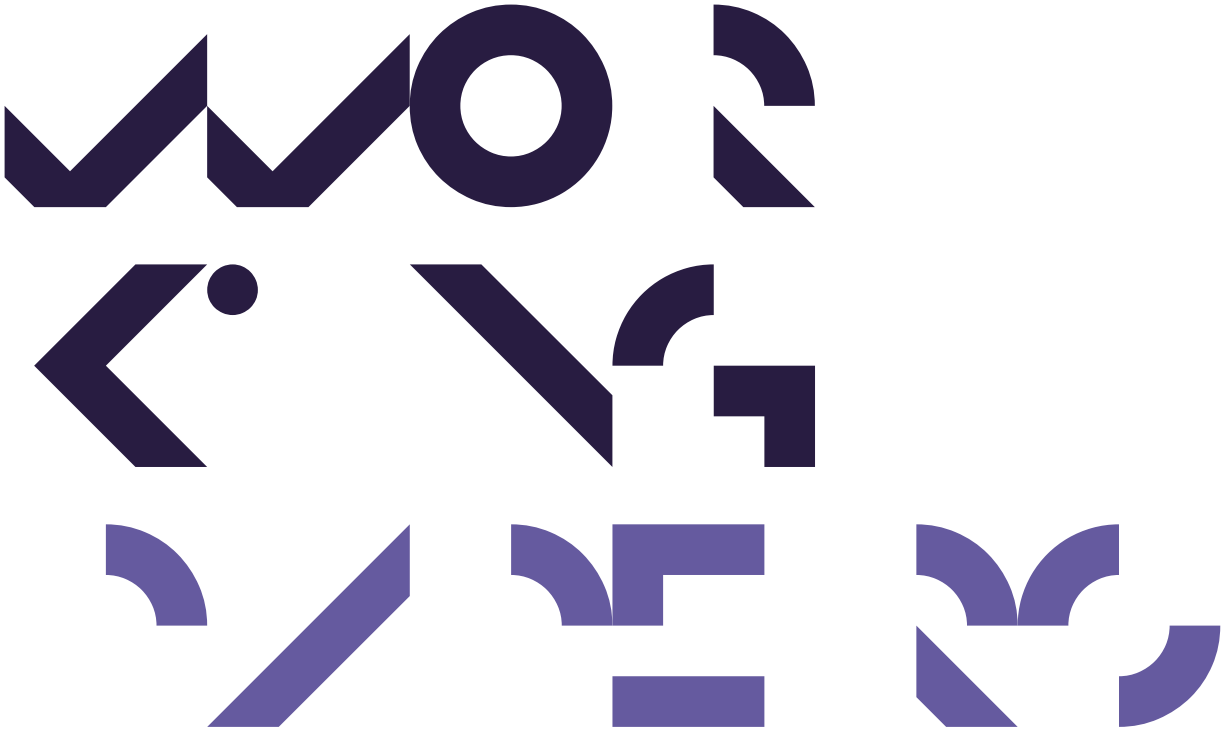


Economics Working Papers

2014-01

Private versus Public Feedback –

The Incentive Effects of Symbolic Awards

Leonie Gerhards and Neele Siemer

AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

Private versus public feedback – The incentive effects of symbolic awards

Leonie Gerhards*[1] and Neele Siemer†[2]

[1]Aarhus University
[2]Goethe-University Frankfurt

August 2014

Abstract

Previous empirical studies have demonstrated the positive incentive effects of private and public recognition of work performance. In our controlled laboratory environment we directly compare these two feedback schemes. We moreover compare their incentive effects across two real-effort tasks that differ with regard to how prestigious the subjects perceive working on them. We find that both feedback schemes similarly lead to performance increases compared to the tasks' respective control treatments. Also the subjects' ability and a positive interim feedback enhance performance. Interestingly, competitive preferences matter only in the less prestigious task. The subjects' gender and overconfidence can in neither of the tasks explain performance.
In a supplementary field experiment conducted at a secondary school we furthermore investigate the incentive effects of different forms of public recognition.

Keywords: Private feedback, public feedback, relative performance, competitive preferences, laboratory experiment, field experiment
JEL classifications: C9, J3, M1

*Corresponding author; Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark; lgerhards@econ.au.dk, Tel.: (+45) 87 16 52 29

†Department of Management and Microeconomics, Goethe-University, Grüneburgplatz 1, D-60323 Frankfurt, Germany; siemer@wiwi.uni-frankfurt.de

1 Introduction

Firms frequently use feedback on relative performance, such as announcing the "employee of the month", as a means to improve performance. The goal of relative performance feedback is to motivate the employees and thereby to increase their effort and productivity. Some companies do this by awarding employees for outstanding performance publicly. Others provide relative performance feedback only privately. A common feature of both feedback schemes is that the employees receive information about their relative standing among their peers. Providing feedback publicly, however, additionally affects the employees' social status vis-à-vis their co-workers, which can give rise to additional positive peer effects (e.g. Kandel and Lazear 1992; Falk and Ichino 2006 or Mas and Moretti 2009). But it might also create adverse effects, for instance on employees who fear negative peer responses when being recognized for their outstanding performance (see e.g. Exline et al. 2004 on so-called sociotropic individuals).

Several field experiments show that bestowing purely symbolic awards leads to sizeable performance increases compared to control conditions in which no performance feedback is provided whatsoever (e.g. Bradler et al. Forthcoming; Delfgaauw et al. 2013; Kosfeld and Neckermann 2011 and Markham et al. 2002). Huberman et al. (2004) even observe that the participants in their lab experiment are willing to incur monetary losses to receive public recognition from their peers.

Also the private revelation of an individual's own performance rank encourages people to work harder in order to improve their relative position among their peers (see for instance the studies by Blanes i Vidal and Nossol 2011; Kuhnen and Tymula 2012; Tran and Zeckhauser 2012; Azmat and Iriberri 2010a; 2010b and Barankay 2012). People are actually willing to go as far as to sabotage the output of others or to artificially increase their own performance to achieve a better position in a ranking (Charness et al. 2013).[1]

With our study we aim at filling the gap between these two strands of literature by comparing the incentive effects of private versus public performance feedback. Subjects in our lab experiment only receive a fixed amount of money for their participation. Their performance is not additionally monetarily incentivised. The sole variation over treatments is the way feedback is provided: In the Control treatment subjects do not receive any relative performance feedback at all. In the Private treatment, they are privately informed about being the best performer in their group (or not). And in the Public treatment, finally, the best performer is additionally publicly praised at the end of the experimental session and receives a symbolic award.

A further novelty of our study is that we investigate how the nature of the work task affects the impact of private and public feedback. One of our tasks resembles a boring, obligatory real-world work task. The other real effort task mirrors a more interesting and prestigious work task. We expect the positive effects of feedback provision on performance to be stronger for the more prestigious task. To the best of our knowledge, no study has yet compared the incentive effects of private and public feedback schemes across tasks. Our study therefore aims to provide firms with information on whether or not the type of task matters when they plan to implement a feedback scheme.

Finally, our lab experimental design allows us to cleanly analyse the incentive effects of interim feedback provision on subjects' performance. It also permits us to elicit the subjects' personal characteristics, such as gender and ability and in particular their task-specific overconfidence and general competitive preferences, and to relate those measures to their behaviour in the different tasks and treatments.

Our main results are as follows. First, in both tasks the subjects provide higher efforts in the Private and Public as compared to the Control treatments. Thus, in both our tasks we could replicate previous findings regarding the productivity enhancing effects of private feedback provision and public recognition, respectively. Second, in neither of our tasks we observe significant performance differences between the Private and Public treatments. Public feedback

[1] Other studies suggest that concerns for status can also have a positive effect on behaving pro-socially in various social dilemmas (see e.g. Gaspart and Seki 2003; Lacetera et al. 2010; Pan and Houser 2011).

in conjunction with a symbolic award given to the best performer thus does not seem to provide additional performance incentives compared to mere private feedback.

Third, high initial ability as well as receiving a positive interim feedback about being the current best performer have a positive motivational effect on subjects' performance in both our tasks. Furthermore, in the more tedious work task the subjects' performance is increasing in their competitive preferences in the Private and Public treatments. In the more enjoyable and more prestigious task, conversely, the subjects' general competitive preferences play a minor role. We present a detailed analysis of these heterogeneous treatment effects in Section 3.4, where we also discuss our results in the light of previous lab and field studies.

Furthermore, we conducted a supplementary field experiment with children at a secondary school. Our findings suggest that the prospect of being given a symbolic award has no additional incentive effect compared to being merely publicly praised. The details of the experimental design and results are presented in Section 4.

2 Experimental setup – Lab experiment

Across subjects we vary the computerized real-effort tasks that subjects work on during the experimental sessions. The first one, the so-called "Slider task", introduced by Gill and Prowse (2012), can be thought of as an experimental translation of an everyday, obligatory task from a real-world setting. In that task, the subjects' computer screens display 48 sliders on straight lines that are labelled from 0 to 100. Using the computer mouse the subjects have to regulate as many sliders as possible to the middle of their respective lines. The current position of each slider is displayed in a small box right next to its line.[2] A slider line is solved correctly when its slider is placed exactly on position 50. We expect the intrinsic motivation to work on this task to be comparatively low.

In the other task, which we will call "Picture task", the subjects solve a set of IQ test matrices from the Raven's progressive matrices test (Raven 1936). They have to complete logic patterns (the pictures) by choosing the appropriate element from a set of six to eight elements that are presented on their computer screen.[3] When a subject selects one of the elements and confirms the choice the next incomplete picture is displayed. Subjects are thus forced to make a decision to proceed to the next picture. The presentation of a new picture, however, does not depend on the correctness of their previous solution. We inform subjects that this task is usually part of a larger IQ test. Our aim, however, was not to scientifically elicit their IQ scores. In fact, we use the number of correctly solved test matrices only as a performance measure that finally determines the group's winner. Our expectation is that subjects enjoy working on that task and will be proud of being the best performer, so that the Picture task ultimately creates a higher incentive to excel than the Slider task. This task represents an interesting and prestigious task in a real firm.

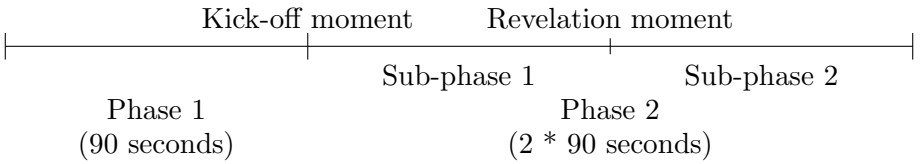


Figure 1: Sequence of events

Each experimental session comprises two phases (see Figure 1). Phase 1 lasts 90 seconds and is identical across treatments. In that phase we ask subjects to provide as much effort as possible in their given task. Subsequently the subjects have to self-assess how many sliders or pictures, respectively, they solved correctly. If their self-assessment is correct, the subjects earn an additional 3 Euro. In our later analysis we use performance in phase 1 as an indicator for

[2] A screen shot of the task can be found in Appendix D.

[3] A sample picture can be found in Appendix E.

an individual's motivation and ability to work on the task.[4] Moreover, the difference between self-assessed and actual performance provides us with a measure of an individual's task-specific overconfidence.

The chronological sequence of phase 2 is deduced from Blanes i Vidal and Nossol's (2011) quasi-experiment. We differentiate, just like them, between a so-called "kick-off" moment, in which the individuals learn that their future performance provides the basis for the coming relative performance feedback, and the "revelation moment", in which the individuals receive feedback about their (interim) performance. Phase 2 thus consists of two sub-phases, each of them lasting 90 seconds, with a short break in-between. At the beginning of phase 2 the subjects receive the instructions for their respective treatment.

In the Control treatment the instructions simply ask the subjects to continue working on their task for a further two times 90 seconds. During the 30 second break between the two sub-phases the computer displays a waiting screen, which asks them to be patient and wait until the experiment continues. This treatment is used to benchmark effort provision over the course of the session in case no feedback is provided.

In the Private and Public treatments the subjects are randomly matched in groups of five and within each of these matching groups the best performing subject in phase 2 (meaning over both sub-phases) is determined as the group's winner.

The instructions for the Private treatment state that subjects are going to be privately informed whether they are their group's winner (or not). This private feedback is provided on their computer screen at the end of phase 2. In case a subject is not a winner, no further rank information is revealed.

In the Public treatment each group's winner is additionally publicly praised. For this the winners have to come to the front of the lab at the end of the session. The non-winning subjects are asked to applaud and to congratulate them. Furthermore, we take photographs of the winners, which are later inserted into "winner certificates" that are posted in a glass case in front of the experimental laboratory for a period of two weeks. A blank version of the certificate is shown on the subjects' screens as part of the instructions at the beginning of phase 2.[5] Our Public treatment is thus particularly inspired by Kosfeld and Neckermann's (2011) field experiment as we also introduce a mere symbolic award which has no further value outside the experimental environment. And more importantly, we also award this prize only to the best performing subject.[6]

In the short break between the two sub-phases we ask the subjects in the Private and Public treatment to self-assess their performance rank within their matching group after the first sub-phase. They receive an additional 3 Euro if their rank estimation is correct. Subsequently, we inform the subjects whether they actually are their group's best performer in sub-phase 1 (or not). Then the second sub-phase starts and the subjects resume to work on their task.

In the Control treatment we similarly insert a break of approximately 30 seconds. All treatments are hence comparable in their temporal structure. The information structures, however, differ. While all treatments are cleanly comparable after the kickoff moment, observed treatment differences between the Control and the competitive treatments after the revelation moment could either be attributable to the prospect of getting (private or public) performance feedback at the end of the session, or to the received interim feedback. Clean treatment comparisons between the Private and Public treatments are, however, still feasible.

At the end of a session the subjects are asked to fill out a short socio-economic questionnaire. Furthermore, we gather data about the subjects' competitive preferences using the trait

[4] We therefore use the terms "phase 1" and "ability phase" interchangeably when referring to this part of the experiment.

[5] A translated version of those certificates can be found in Appendix F.

[6] This allows us to focus on the incentive effects of *positive* feedback for outstanding performance. Alternative approaches are found in e.g. Tran and Zeckhauser (2012), who perform a field experiment in an educational context and provide the students with performance feedback over the full ranking scale. They observe that in particular high ability students increase their performance when private feedback is provided. But in the public feedback treatment, both high ability students who strive to be the best as well as low ability students who aim at preventing to be publicly ranked on the last place increase their efforts.

competitive scale developed by Brown et al. (1998)[7] and their personality traits as measured by a short version of the Big Five questionnaire (see e.g. Goldberg 1992 and Dohmen et al. 2008).

The experiment was conducted at the FLEX laboratory of the Goethe-University Frankfurt. We invited students from various fields of study using ORSEE (Greiner 2004). We aimed at having fairly equal shares of males and females in each of the sessions and the result was 52 percent female subjects. Each subject participated in only one of the treatments and worked on only one task to which he or she was randomly assigned. A total of 279 subjects participated in our experiment, 99 of them in the Control treatments (50 in the Slider, 49 in the Picture task), 95 in the Private treatments (45 in the Slider, 50 in the Picture task) and 85 in the Public treatments (40 in the Slider, 45 in the Picture task). In all our treatments we paid the subjects a fixed amount of money (8 Euro) at the beginning of the session and did not provide any additional monetary performance incentives. In both tasks only the number of correctly solved sliders or pictures, respectively, determined the subjects' performance – and in the Private and Public treatments also the winner within each matching group. All decisions were made on a computer screen using z-Tree (Fischbacher 2007). Sessions lasted approximately one hour, during which the subjects earned an average of 9 Euro.

3 Results

3.1 Perception of the tasks

To check whether the subjects perceive the experimental tasks as intended, we ask them to rate various dimensions of their respective task in the final questionnaire. Figure 2 displays their answers (for the exact wording of the questions, answer options and procedures see Appendix A). The results show first that the Picture task is indeed perceived as more enjoyable and more challenging than the Slider task (results from two-sided Mann-Whitney ranksum tests: $p < 0.001$ and $p < 0.001$, respectively). Second, the subjects do not indicate to provide higher efforts in one of the tasks ($p = 0.898$), nor do they report significantly different levels of exhaustion ($p = 0.900$). And finally, they state a higher desire to win in the Picture task ($p = 0.099$). This makes us confident that our experimental task manipulation works as anticipated.

We furthermore do not observe significant differences in male and female performance in the ability phase of the two tasks (two-sided Mann-Whitney ranksum test result: $p = 0.938$ for equivalence of men's and women's number of correctly solved sliders in the Slider task, $p = 0.551$ for the number of solved pictures in the Picture task). Both our tasks thus seem to be gender neutral as they do not favour one of the two sexes.

Moreover, given task we observe in neither of the treatments significant differences in the distribution of subjects' competitive preferences as measured by Brown et al.'s (1998) trait competitive scale (two-sided Mann-Whitney ranksum test results for the Slider task: Control vs. Private: $p = 0.914$, Control vs. Public: $p = 0.341$, Private vs. Public: $p = 0.375$; for the Picture task: Control vs. Private: $p = 0.604$, Control vs. Public: $p = 0.606$, Private vs. Public: $p = 0.869$). Given that the subjects are randomly allocated to the treatments this finding reassures us that the elicited preferences are not biased by the treatments in which subjects previously participated. They can therefore be used as reliable information about the subjects' general competitive preferences.[8]

3.2 Slider task

The beginning of sub-phase 1 constitutes the kick-off moment. At that point in time the subjects are informed about their treatment-specific feedback rule. Performance in sub-phase 1 thus reflects the pure incentive effect of that announcement. Receiving the interim feedback about

[7] A copy of Brown et al.'s (1998) trait competitive scale can be found in Appendix B.

[8] To further check the reliability of this competitive preference measure we asked subjects whether they practice any sports in which they take part in competitions. The elicited general competitive preferences and the answers to the competitive sports question are significantly correlated ($\beta = 0.159, p = 0.008$).

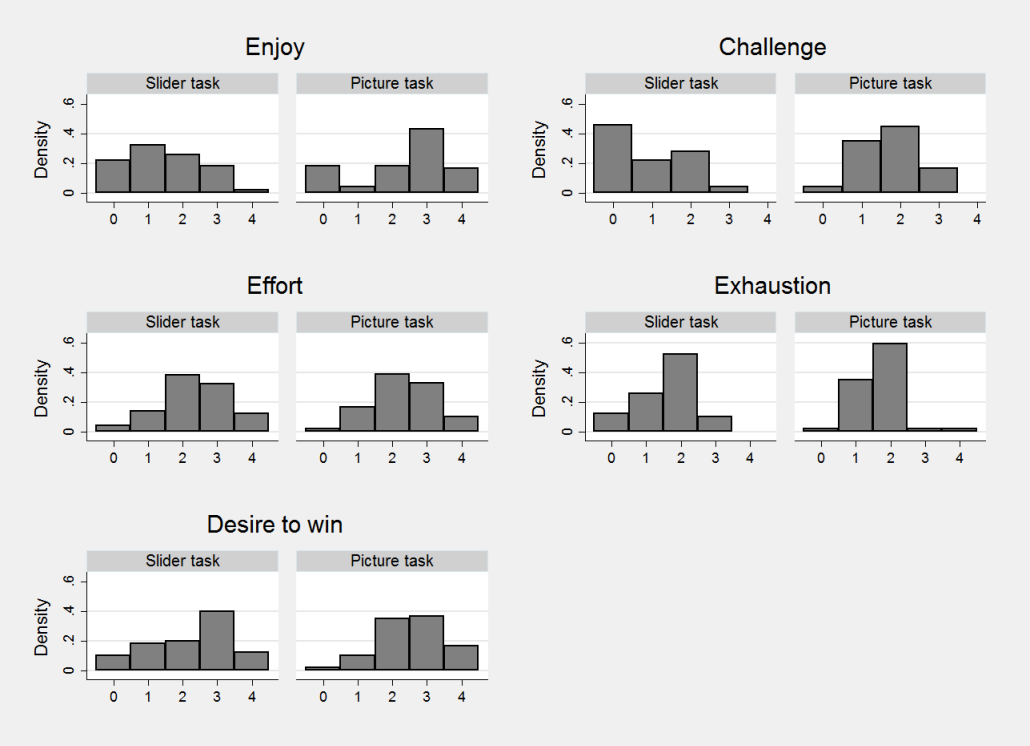


Figure 2: Perception of the tasks

being the group's best performer in sub-phase 1 (or not) constitutes the revelation moment. The subjects' performance response to that feedback is recorded in sub-phase 2.

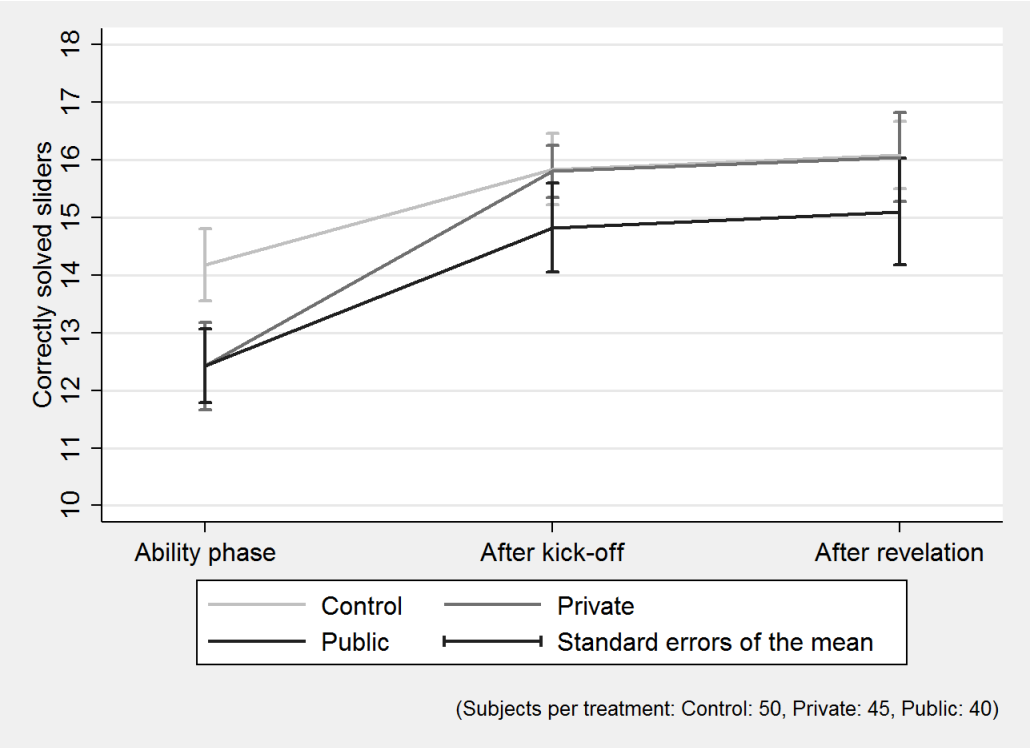


Figure 3: Performance over time (Slider task)

Figure 3 and Table C.1 in Appendix C display the average numbers of correctly solved sliders over (sub-)phases by treatments. It immediately becomes evident that subjects exhibit learning effects in all treatments as the average numbers of correctly solved sliders increase

over time. Moreover, we observe significant treatment differences: For both the Private and the Public treatment the absolute performance increase between the first sub-phase of phase 2 (i.e. after the kick-off moment) and the ability phase is significantly larger than in the Control treatment (two-sided Mann-Whitney ranksum tests: $p = 0.016$ and $p = 0.032$, respectively). The performance increases in the Private and Public treatment, however, are virtually identical ($p = 0.972$). These findings translate to the second sub-phase. When we compare the absolute difference between performance in the second sub-phase (i.e. after the revelation moment) and performance in the ability phase, we again find significant differences between the Private and the Control treatment ($p = 0.035$) as well as between the Public and the Control treatment ($p = 0.076$).[9] But the difference in performance between the two competitive treatments, Private and Public, is not significant ($p = 0.624$).[10]

Result 1 *The subjects' average performance increases significantly more in the Private and Public treatments (compared to the Control treatment) once future performance feedback is announced and interim feedback is provided. The performance improvements in the two competitive treatments, however, are not statistically different from each other.*

Note that Result 1 cannot simply be explained by a potential limit to the more able Control treatment subjects' possible improvements in performance from phase 1 to phase 2. The behavioural difference across treatments is more fundamental. Two-sided Fisher exact tests reveal that the fraction of subjects who do not improve their performance after the kick-off moment is significantly larger in the Control treatment as compared to the Private and Public treatments (Control vs. Private: 38% vs. 18%, $p = 0.040$, Control vs. Public: 38% vs. 18%, $p = 0.038$). The fraction of subjects who react with a performance increase to the announcement of future feedback, however, is virtually identical in the Private and Public treatment ($p = 1.0$).

In the following we analyse heterogeneous treatment effects with respect to the subjects' personal characteristics. For this purpose we run a series of OLS regressions. The results are presented in Table 1. Note that we standardise all continuous explanatory variables as well as the dependent variable in order to be able to compare the regression coefficients across tasks later. In the first two specifications we regress performance after the kick-off moment, i.e. the number of correctly solved sliders in the respective sub-phase, on the subjects' competitive preferences and their number of correctly solved sliders in the ability phase. We build interaction terms of those variables and the treatment dummies and furthermore control for the subjects' gender, their personal characteristics elicited in the Big Five questionnaire and the subjects' overconfidence. This is defined as the difference between their stated number of correctly solved sliders in the ability phase and the actual number of correctly solved sliders. The first specification compares the performance across all three treatments, whereas the model in column 2 focuses on the comparison between the Private and Public treatment.

From the first two specifications we deduce that highly competitive subjects put in considerably more effort in the competitive treatments than in the Control treatment. Moreover, the impact of competitive preferences on performance is even higher in the Public treatment as compared to the Private treatment. Also ability has a significantly positive effect on performance in all treatments, though its magnitude differs across treatments. Its effect is highest in the Control, medium in the Public and lowest in the Private treatment.

In the right part of the table we run analogue regressions using subjects' (standardised) performance after the revelation moment as dependent variable. In the last column we add a dummy variable for the best performing subject in the first sub-phase of phase 2 and compare the effects of this positive interim feedback on performance between the Private and Public treatment, while controlling for the impact of expecting to be the best performer.

[9] As discussed in Section 2 above, this could either be attributable to the previous announcement of the respective feedback scheme or to the interim feedback given in the revelation moment.

[10] Note that we opted for the analysis of the differences between sub-phase performance and ability phase performance instead of the raw numbers of sub-phase performances, as it allows us to control for subject-specific ability fixed effects captured in their performance in phase 1.

Table 1: Determinants of the number of solved sliders

	After kick-off		After revelation	
	All treatments	Private vs. Public	All treatments	Private vs. Public
Competitive preferences	0.005	0.156*	0.089	0.349***
	(0.092)	(0.081)	(0.095)	(0.083)
Competitive preferences x Private	0.159		0.310**	
	(0.117)		(0.130)	
Competitive preferences x Public	0.449***	0.308**	0.257	-0.177
	(0.142)	(0.147)	(0.161)	(0.159)
Ability	0.856***	0.406***	0.544***	0.064
	(0.109)	(0.121)	(0.128)	(0.132)
Ability x Private	-0.435***		-0.309*	
	(0.157)		(0.173)	
Ability x Public	-0.094	0.356**	0.079	0.327
	(0.151)	(0.166)	(0.225)	(0.260)
Male	-0.063	-0.032	0.118	0.099
	(0.151)	(0.223)	(0.163)	(0.247)
Overconfidence	-0.016	0.001	-0.089	-0.155*
	(0.065)	(0.091)	(0.056)	(0.082)
Agreeableness	0.014	-0.053	-0.068	-0.214
	(0.094)	(0.122)	(0.109)	(0.129)
Conscientiousness	-0.054	-0.023	0.021	0.005
	(0.104)	(0.135)	(0.123)	(0.152)
Extraversion	-0.052	-0.068	-0.160*	-0.128
	(0.079)	(0.120)	(0.093)	(0.141)
Neuroticism	0.073	0.093	0.025	0.074
	(0.072)	(0.101)	(0.087)	(0.126)
Openness	0.044	0.069	0.058	0.111
	(0.071)	(0.099)	(0.068)	(0.087)
Winner after kick-off				0.771***
				(0.264)
Winner after kick-off x Public				0.340
				(0.455)
Expected to be winner after kick-off				-0.080
				(0.243)
Private	0.252**		0.166	
	(0.124)		(0.156)	
Public	0.107	-0.166	0.055	-0.245
	(0.184)	(0.196)	(0.192)	(0.256)
Constant	-0.096	0.150	-0.136	-0.125
	(0.117)	(0.131)	(0.115)	(0.177)
Observations	135	85	135	85
R^2	0.51	0.40	0.38	0.43

All continuous independent variables and the respective dependent variables are standardised. Robust standard errors are reported in parentheses: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

The regression results from column 3 resemble the findings from the first sub-phase (see column 1). Competitive subjects exert higher efforts in the Private and Public treatments as compared to the Control treatment (the coefficient for the respective interaction term for the Public treatment is $p = 0.112$ and thus only borderline insignificant). Also the subjects' ability affects their performance positively in all treatments. When we focus on the Private and Public treatments in the regression in the last column, we find that the subjects' competitive preferences still have a positive impact on performance in the Private treatment. Furthermore, these treatments do not differ in this respect as indicated by the insignificant coefficient of the interaction term. The coefficient for ability, on the other hand, becomes insignificant. Instead, it is now the positive interim feedback (i.e. receiving the information to be the winner after the first sub-phase) that has a significantly positive influence on performance in the Private treatment – even when controlling for the expectation to be the best performer in the first sub-phase. The corresponding insignificant coefficient of the interaction term furthermore indicates that its incentive effect is not significantly different between the Private and Public treatment.

The gender dummy and our measure for the subjects' overconfidence are not significant in any of the four specifications of Table 1. Also the personality traits elicited in the Big Five questionnaire have no robust predictive power across treatments and sub-phases. The main findings are summarized below:

Result 2 *The overall positive treatment differences between the two competitive treatments and the Control treatment in the Slider task are particularly driven by subjects with high competitive preferences and those who receive a positive interim feedback, which in turn is related to the subjects' ability.*

Note that these findings are independent of the inclusion of the gender dummy, the overconfidence measure and the Big 5 characteristics in the OLS specifications. The corresponding regressions, in which we dropped those variables, are displayed in Table C.2 in Appendix C.

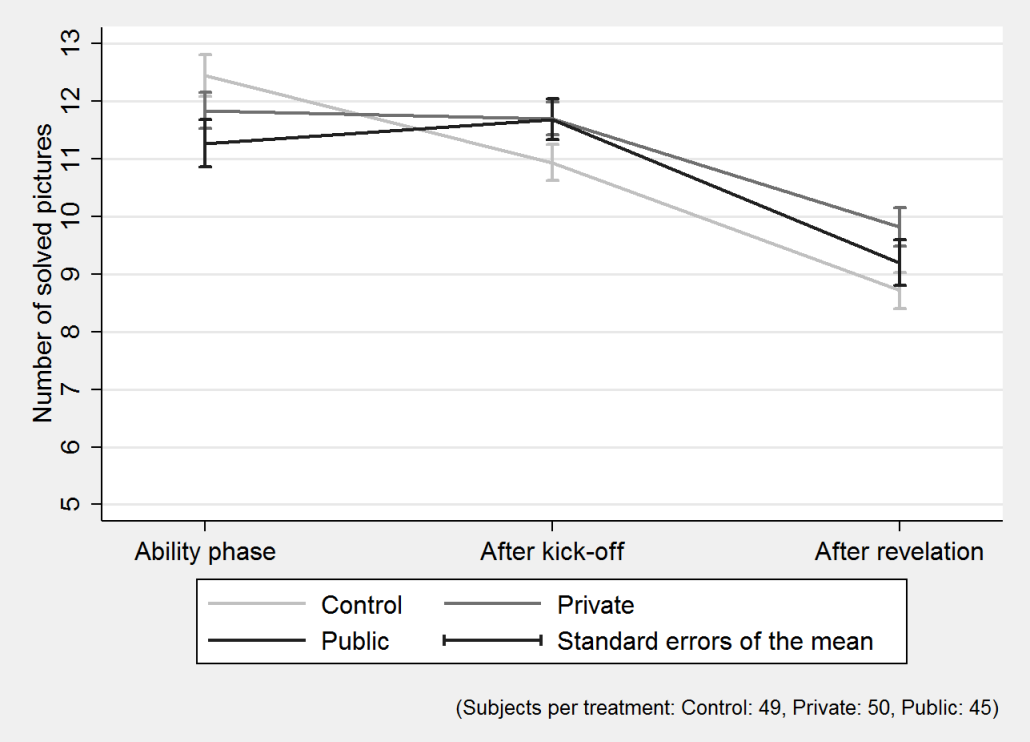


Figure 4: Performance over time (Picture task)

3.3 Picture task

Figure 4 and Table C.3 in Appendix C present the average total number of solved pictures over time. Although only the number of correctly solved pictures determined each groups' winner in that task, we decide to use the total number of completed pictures (both correctly and wrongly) as a subject's measure of performance. The reason is that subjects in the Picture task cannot know for sure if they solved a picture correctly or not when entering their choice – in contrast to the Slider task, where they can immediately see if they regulated a slider to the middle of its line. It may therefore well be possible that subjects also exerted high effort on a picture that was finally solved incorrectly. The fact that subjects who participated in the Picture task report equally high levels of provided effort and exhaustion as those who worked on the Slider task (see Section 3.1), supports the assumption that the subjects actually tried to solve the majority of the pictures and did not only click through them. Besides that, the general pattern of results remains unchanged when we focus only on the number of correctly solved pictures in the statistical analysis. The corresponding descriptive statistics and regression tables are reported in Tables C.4 and Table C.5 Appendix C.

It is evident that the subjects' measured performance decreases over time in all treatments. This effect is attributable to the fact that the average level of difficulty of the pictures displayed in phase 2 is higher than in phase 1.[11] But despite the observed general negative trend in measured performance we are able to study treatment differences, particularly as the pictures are for all subjects presented in the same sequence in all treatments and all (sub-)phases.[12] Like we did for the Slider task we again analyse the subjects' absolute performance change after the kick-off as well as after the revelation moment compared to their performance in the ability phase. Two-sided Mann-Whitney ranksum tests reveal that the measured productivity after the kick-off moment decreases significantly less in the Private as compared to the Control treatment ($p = 0.004$). The same is true when comparing the Public to the Control treatment ($p < 0.001$). The performance changes in the Private and Public treatments, however, do not differ significantly ($p = 0.218$). This pattern translates to the performance after the revelation moment. The subjects' performance decreases significantly less in the Private and Public treatments than in the Control treatment (two-sided Mann-Whitney ranksum test results: $p = 0.001$ and $p < 0.001$, respectively). The respective numbers for the Private and Public treatments are again virtually identical ($p = 0.746$).

Using the OLS regressions presented in Table 2 we aim once more at eliciting heterogeneous treatment effects, just as we did for the Slider task. The general structure of Table 2 is comparable to the one of Table 1. The findings differ, however. The subjects' competitive preferences are found to have virtually no impact on performance in the Picture task. In all treatments, ability positively affects performance after the kick-off moment. And the last specification (column 4) shows that instead of ability, receiving a positive interim feedback about being the best performer in sub-phase 1 has significantly positively affects performance in the Public treatment.

Just like in the corresponding regressions for the slider task (Table 1) the gender dummy as well as the overconfidence measure and the personality factors elicited in the Big Five questionnaire have no robust predictive power across treatments and sub-phases. We summarize our findings below:

Result 3 *The subjects' competitive preferences that explained the general treatment effects in the Slider task do not play any significant role in the Picture task. Instead, subjects' ability, and in the Public treatment particularly receiving a positive interim feedback after the first sub-phase, affect performance significantly positively.*

[11] We tried to avoid this effect when programming the order in which the pictures are displayed compared to the original test. Our aim was to present pictures of approximately similar difficulty in phase 1, and the two sub-phases of phase 2, respectively. However, our randomization of pictures evidently did not suffice to counter the effect of a seemingly decreasing productivity.

[12] In each sub-phase we provide the subjects with a predefined set of pictures, irrespective of how many pictures they solved in the previous sub-phase.

Table 2: Determinants of the number of solved pictures

	After kick-off		After revelation	
	All treatments	Private vs. Public	All treatments	Private vs. Public
Competitive preferences	-0.001	0.115	-0.053	0.061
	(0.128)	(0.110)	(0.105)	(0.097)
Competitive preferences x Private	0.132		0.108	
	(0.177)		(0.144)	
Competitive preferences x Public	-0.171	-0.305	0.066	-0.111
	(0.207)	(0.185)	(0.176)	(0.160)
Ability	0.613***	0.489***	0.307***	0.535***
	(0.152)	(0.108)	(0.102)	(0.120)
Ability x Private	-0.126		0.128	
	(0.185)		(0.153)	
Ability x Public	-0.350	-0.237	-0.329	-0.700***
	(0.243)	(0.212)	(0.202)	(0.214)
Male	-0.337**	-0.275	0.044	0.162
	(0.159)	(0.191)	(0.142)	(0.182)
Overconfidence	-0.037	-0.020	-0.027	0.118
	(0.085)	(0.097)	(0.068)	(0.099)
Agreeableness	0.163	0.032	0.024	0.040
	(0.100)	(0.126)	(0.093)	(0.144)
Conscientiousness	-0.215**	-0.137	-0.041	0.110
	(0.108)	(0.139)	(0.096)	(0.123)
Extraversion	-0.022	-0.134	0.078	-0.019
	(0.097)	(0.126)	(0.097)	(0.123)
Neuroticism	-0.036	0.028	0.114	0.138
	(0.075)	(0.082)	(0.073)	(0.084)
Openness	-0.092	-0.105	0.002	0.074
	(0.073)	(0.097)	(0.077)	(0.093)
Winner after kick-off				-0.025
				(0.276)
Winner after kick-off x Public				0.966**
				(0.407)
Expected to be winner after kick-off				0.516**
				(0.218)
Private	0.369**		0.290**	
	(0.186)		(0.146)	
Public	0.412*	0.038	-0.128	-0.840***
	(0.221)	(0.210)	(0.214)	(0.282)
Constant	-0.120	0.213	-0.107	0.059
	(0.148)	(0.146)	(0.117)	(0.179)
Controlling for mistakes	Yes	Yes	Yes	Yes
Observations	144	95	144	95
R^2	0.35	0.36	0.43	0.50

All continuous independent variables and the respective dependent variables are standardised. Robust standard errors are reported in parentheses: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Like in the Slider task, these findings do not depend on the inclusion of the gender dummy, the overconfidence measure and the Big 5 characteristics in the OLS specifications. The corresponding regressions, in which those variables are left out, are presented in Table C.6 in Appendix C.

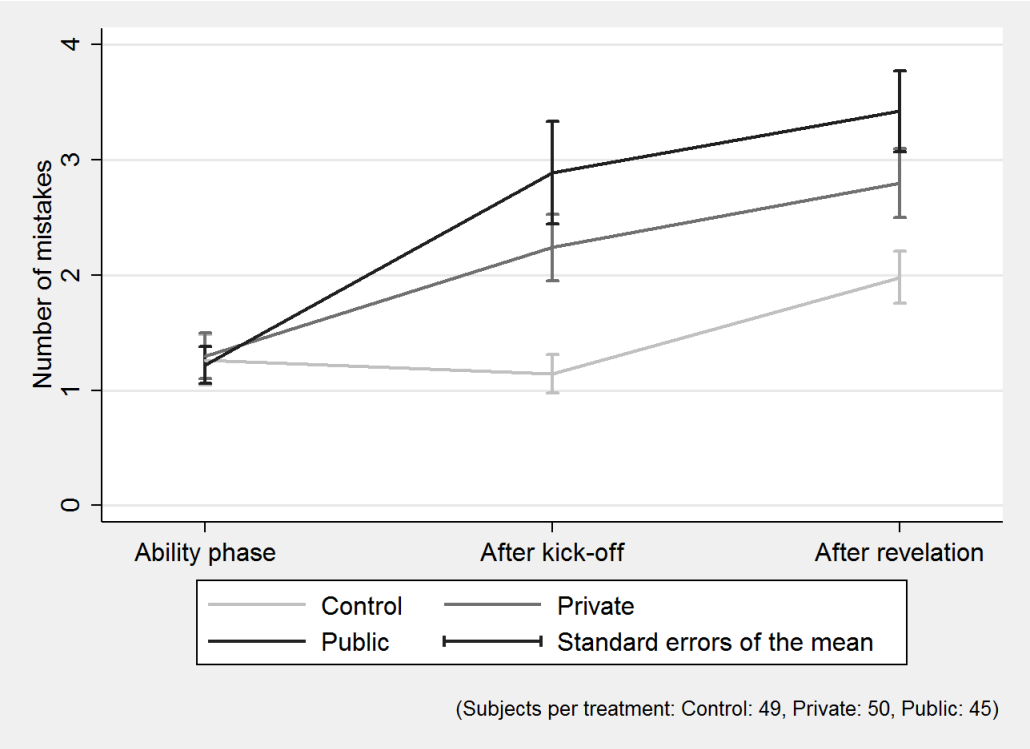


Figure 5: Mistakes over time (Picture task)

A further advantage of the Picture task is that we can measure performance over treatments not merely in quantity, but also as regards quality. Figure 5 depicts the average number of wrongly solved pictures across treatments over the three (sub-)phases of the experiment. As is evident the number of mistakes in the ability phase is virtually similar across treatments. After the kick-off moment as well as after the revelation moment, however, the subjects make significantly more mistakes in the competitive treatments as compared to the Control treatment (two-sided Mann-Whitney ranksum test results: after the kick-off moment: Private vs. Control: $p = 0.005$, Public vs. Control: $p < 0.001$, after the revelation moment: Private vs. Control: $p = 0.047$, Public vs. Control: $p = 0.002$). At the same time the subjects from the Private and Public treatments evaluate the Picture task as a greater challenge compared to those from the Control treatment (two-sided Mann-Whitney ranksum test results: Private vs. Control: $p = 0.015$, Public vs. Control: $p = 0.006$, Private vs. Public: $p = 0.717$). These observations lead to the following result:

Result 4 *The competitive environments of the Private and Public treatments put subjects under considerable pressure to perform, which leads to increasing mistakes as compared to the Control treatment.*

This psychological phenomenon is also known as "choking under pressure" and has already been documented in various lab and field contexts (see e.g. Baumeister 1984 or Dandy et al. 2001 for choking under pressure in competitions). Interestingly, however, just like the total numbers of pictures solved, also the numbers of mistakes made in the Private and Public treatment, are neither after the kick-off moment, nor after the revelation moment significantly different (two-sided Mann-Whitney ranksum test results: $p = 0.502$ and $p = 0.202$, respectively).

3.4 Discussion of our findings in the light of previous studies

The overall treatment effects are remarkably similar in the Slider task and the more enjoyable and prestigious Picture task. In both tasks subjects put in significantly higher efforts in the Private and Public compared to the Control treatment, both after the kick-off and after the revelation moment. Given task the performance in our Private and Public treatments does not differ substantially, however. This suggests that private as well as public feedback mechanisms provide similar performance enhancing effects.

Our findings on positive effects of private feedback provision from both tasks are thus very much in line with results from field studies conducted at firms (e.g. Blanes i Vidal and Nossol 2011) or in educational contexts (e.g. Azmat and Iriberri 2010a and Tran and Zeckhauser 2012) as well as with previous lab experimental results (e.g. Kuhnen and Tymula 2012 and Azmat and Iriberri 2010b). This is in particular interesting as those studies analyse situations in which individuals are provided with performance rankings relative to their peers (or at least relative to the average performance of their peers), whereas the subjects in our experiment only receive information on whether they are the winner in their group or not.

Furthermore, although being publicly praised as a winner in a lab experiment might appear artificial, the Public treatments in our tasks generate similar positive performance effects as public recognition of employees studied in natural field settings (e.g. Kosfeld and Neckermann 2011, Tran and Zeckhauser 2012 and Delfgaauw et al. 2013).

The Big Five measures and our measure for overconfidence are not able to significantly and robustly explain the subjects' performance in either of the two tasks. Interestingly, the same is true for the subjects' gender. This is in line with the studies of Blanes i Vidal and Nossol (2011) and Azmat and Iriberri (2010a), who also do not find that males and females react differently to the competitive environments that they analyse. However, it is in contrast to the studies of Azmat and Iriberri (2010b), Kuhnen and Tymula (2012), Barankay (2012) and Delfgaauw et al. (2013), who observe that males respond more strongly to private or public feedback provision.[13] We conjecture that this pattern in our findings can be explained by the gender-neutral nature of our tasks (see section 3.1 above).

It is interesting to note that the subjects' competitive preferences only act as a significant predictor of performance in the competitive treatments in the Slider task, both, after the kick-off as well as after the revelation moment. This suggests that once feedback schemes are installed in a rather boring work task, predominantly subjects who want to satisfy their preference for competition react positively to the implemented scheme. In a more interesting work task, conversely, competition-affine and competition-averse subjects alike react to the implemented treatments.

In both our tasks the subjects' ability had a strong and comparable influence on the observed treatment effects, which replicates findings from previous studies (e.g. Kosfeld and Neckermann 2011). More skilled individuals react more positively on competitive environments.

However, in both tasks the ability's impact vanishes once one controls for the received interim feedback in the revelation moment. In the Picture task only the interim winners in the Public treatment significantly improve their performance in the last sub-phase. In the Slider task, conversely, interim winners in both the Private and Public treatments increase their performance significantly. In this regard our results are very much in line with Azmat and Iriberri (2010b) who similarly find that positive interim feedback increases performance. But also a negative feedback leads to performance increases in their lab experiment. In general, studies on the incentive effects of interim feedback come to mixed results. In Barankay's (2012) field experiment, for instance, positive feedback has no effect on performance whereas negative feedback has a significant negative effect on observed performance. Bradler et al. (Forthcoming) find that positive feedback (in their case receiving an award) has virtually no effect on performance whereas not receiving an award encourages subjects to work harder. Hoogveld and Zubanov (2014) observe

[13] But note that also Delfgaauw et al. (2013) do not find that male managers generally react more positively on the implemented competitive treatment. It is rather the combination of a male manager and a high proportion of male employees that increases performance in their competitive treatments.

that neither recipients, nor non-recipients of unannounced public feedback significantly change their subsequent performance in a university classroom context. And Kuhnen and Tymula's (2012) results, finally, suggest that a better than expected feedback decreases individuals' performance, whereas a worse than expected feedback increases performance. It is up to future research to identify the personal or environmental factors that drive these contradicting results.

4 Field study

Besides studying the incentive effect of receiving performance feedback publicly as compared to privately we are furthermore interested in disentangling the effect of being publicly bestowed with a symbolic award compared to merely receiving public feedback. For this purpose we conducted an additional field experiment with 78 pupils (41 boys, 38 girls) at a German secondary school. The children were all in the sixth grade, between 11 and 12 years old, and belonged to three different classes. They were not aware of the fact that they participated in an experiment and did not receive any monetary or other compensation. As our measure of performance we use the total distance that each child was able to hop on one leg without a break.

The study was conducted during two sports lessons that were given on different days. On day 1 the children from each class were simply asked to hop on one leg as far as they could. We did not announce that we would come back on a later day and the children did not receive any information about future performance evaluations or treatments.[14] On the second day (two weeks later) each class was randomly assigned to one of the following treatments: No Competition, Competition and Award, which were explained to them verbally. In the No Competition treatment the children were simply asked to hop again as far as they could. This group provided a control group for non-observable factors that might cause differences in performance between day 1 and day 2. Children in the Competition and in the Award treatment, on the other hand, were informed that the child who improved his or her performance (in absolute terms) the most as compared to the first day would be publicly announced in front of all the other children. In the Award treatment the children were additionally told that the child, who improved his or her performance the most would receive a certificate. This was shown to the class before measuring the children's performance on day 2. A translated version of the certificate can be found in Appendix G.

Table 3: Summary of performance and performance changes from day 1 to day 2

Treatment	Distance hopped on day 1	Distance hopped on day 2	Difference between day 1 and day 2
No Competition	112.782	97.618	-15.164
	(36.685)	(37.690)	(23.494)
Competition	89,762	98.304	8.542
	(45.467)	(43.391)	(44.683)
Award	125,177	150.885	25.708
	(57.848)	(103.699)	(61.068)

28 children participated in the No Competition, 24 in the Competition and 26 in the Award treatment. The mean values of distances hopped on day 1 to day 2 and their difference are denoted in meters. Standard deviations are given in parentheses.

Table 3 summarises the measured performance on days 1 and 2 as well as the performance changes between those two days. Interestingly, in the No Competition treatment the distance hopped on the second day decreases by 12% as compared to the first day. Conversely, the children increase their average performance by 26% in the Competition treatment and by 20% in

[14]Note that the task is gender neutral in the sense that girls and boys exhibit no significant differences in their abilities to hop on the first day ($p = 0.737$).

the Award treatment. The children's performance thus increases significantly more in the Competition than in the No Competition treatment (result from two-sided Mann-Whitney ranksum test: $p = 0.013$). Also the difference between the No Competition and the Award treatment is statistically significant ($p < 0.001$). However, we do not observe a significant difference in performance changes from day 1 to day 2 between the Competition and the Award treatment ($p = 0.449$).

We thus conclude that the mere prospect of being publicly announced as the winner has a significantly positive effect on the children's performance on day 2. An additional symbolic award, however, provides no supplementary incentive effect.

5 Conclusion

In many fields such as sports, arts and organisations, competitions and awarding prizes are a means to reward outstanding performance. One strand of literature shows that purely symbolic awards generate a positive incentive effect (e.g. Kosfeld and Neckermann 2011). Another strand demonstrates that even a merely private feedback provision about the individuals' relative positions has performance enhancing effects (e.g. Blanes i Vidal and Nossol 2011).

Results from both our lab and field experiments consistently point to the conclusion that the method used to reward the winner in a competitive environment does not matter. Providing a non-material award for the winner in a competitive environment does not have an additional effect on performance compared to merely announcing the winner publicly, which in turn does not generate significantly higher incentives than announcing the winner privately. But a common characteristic of these feedback schemes is that they give rise to significantly higher performance compared to conditions in which no feedback is given.

If a firm plans to use relative feedback provision solely as an incentive device, our findings therefore suggest that the company can refrain from organising costly appraisal ceremonies and instead provide this feedback only privately.[15] Furthermore, our lab experiment shows that performance feedback affects people differently, depending on their personal characteristics and the task in which the feedback is provided. High ability subjects respond positively to the competitive treatments in both our tasks. In the less enjoyable and less challenging Slider task the positive treatment effects of competitive environments are furthermore driven by subjects with a particularly high preference for competitions. Firms should thus take into account that predominantly competitive workers might react to the introduction of (private or public) feedback if it is provided in standard, obligatory work tasks. On the other hand, if the feedback is provided on more interesting and more prestigious tasks, the competitive preferences seem to play a minor role and thus a much larger part of the workforce can be expected to respond to the feedback scheme.

Acknowledgements

This work was financially supported by the research grant *"Kleinere Projekte zur Frauen- und Genderforschung"*, which we gratefully acknowledge.

References

Azmat, G., Iriberri, N., 2010a. The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students. Journal of Public Economics 94 (7–8), 435–452.

[15] Obviously, public appraisal ceremonies come with various additional features which have the potential to yield positive outcomes for firms that we did not focus on in our present study. For example, they can be used as a means to remind the workforce of the organisation's values and culture as well as on general objectives and behavioural norms (see e.g. Baron and Kreps 1999, Chapter 10).

Azmat, G., Iriberri, N., 2010b. The Provision of Relative Performance Feedback Information: An Experimental Analysis of Performance and Happiness. Working Papers (Universitat Pompeu Fabra. Departamento de Economía y Empresa) (1216).

Barankay, I., 2012. Rank Incentives - Evidence from a Randomized Workplace Experiment. Working Paper (The Wharton School, University of Pennsylvania).

Baron, J. N., Kreps, D. M., 1999. Strategic Human Resources: Frameworks for General Managers. Vol. 149. John Wiley New York.

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.

Blanes i Vidal, J., Nossol, M., 2011. Tournaments Without Prizes: Evidence from Personnel Records. Management Science 57 (10), 1721–1736.

Bradler, C., Dur, R., Neckermann, S., Non, A., Forthcoming. Employee Recognition and Performance: A Field Experiment. Journal of Labor Economics.

Brown, S. P., Cron, W. L., Slocum Jr, J. W., 1998. Effects of Trait Competitiveness and Perceived Intraorganizational Competition on Salesperson Goal Setting and Performance. The Journal of Marketing, 88–98.

Charness, G., Masclet, D., Villeval, M. C., 2013. The Dark Side of Competition for Status. Management Science 60 (1), 38–55.

Dandy, J., Brewer, N., Tottman, R., 2001. Self-consciousness and Performance Decrements Within a Sporting Context. The Journal of Social Psychology 141 (1), 150–152.

Delfgaauw, J., Dur, R., Sol, J., Verbeke, W., 2013. Tournament Incentives in the Field: Gender Differences in the Workplace. Journal of Labor Economics 31 (2), 305–326.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2008. Representative Trust and Reciprocity: Prevalence and Determinants. Economic Inquiry 46 (1), 84–90.

Exline, J. J., Single, P. B., Lobel, M., Geyer, A. L., 2004. Glowing Praise and the Envious Gaze: Social Dilemmas Surrounding the Public Recognition of Achievement. Basic and Applied Social Psychology 26 (2-3), 119–130.

Falk, A., Ichino, A., 2006. Clean Evidence on Peer Effects. Journal of Labor Economics 24 (1), 39–57.

Fischbacher, U., 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. Experimental Economics 10 (2), 171–178.

Gaspart, F., Seki, E., 2003. Cooperation, Status Seeking and Competitive Behaviour: Theory and Evidence. Journal of Economic Behavior & Organization 51 (1), 51–77.

Gill, D., Prowse, V., 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. American Economic Review 102 (1), 469–503.

Goldberg, L. R., 1992. The Development of Markers for the Big-Five Factor Structure. Psychological Assessment 4 (1), 26–42.

Greiner, B., 2004. An Online Recruitment System ORSEE 2.0 - A Guideline for the Organization of Experiments in Economics. Working Paper Series in Economics 10 (23), 63–104.

Helmreich, R. L., Spence, J. T., 1978. The Work and Family Orientation Questionnaire: An Objective Instrument to Assess Components of Achievement Motivation and Attitudes Toward Family and Career. American Psycholog. Ass., Journal Suppl. Abstract Service.

Hoogveld, N., Zubanov, N., 2014. The Power of (no) Recognition: Experimental Evidence From the University Classroom. Working Paper.

Huberman, B. A., Loch, C. H., Önçüler, A., 2004. Status as a Valued Resource. Social Psychology Quarterly 67 (1), 103–114.

Kandel, E., Lazear, E. P., 1992. Peer Pressure and Partnerships. Journal of Political Economy (4), 801–817.

Kosfeld, M., Neckermann, S., 2011. Getting More Work for Nothing? Symbolic Awards and Worker Performance. American Economic Journal: Microeconomics 3 (3), 86–99.

Kuhnen, C. M., Tymula, A., 2012. Feedback, Self-Esteem, and Performance in Organizations. Management Science 58 (1), 94–113.

Lacetera, N., Macis, Mario, 2010. Social Image Concerns and Prosocial Behavior: Field Evidence from a Nonlinear Incentive Scheme. Journal of Economic Behavior & Organization 76 (2), 225–237.

Markham, S. E., Scott, D., McKee, G., 2002. Recognizing Good Attendance: A Longitudinal Quasi-experimental Field Study. Personnel Psychology 55, 639–660.

Mas, A., Moretti, E., 2009. Peers at Work. The American Economic Review 99 (1), 112–145.

Pan, X. S., Houser, D., 2011. Competition for Trophies Triggers Male Generosity. PloS one 6 (4), e18050.

Raven, J. C., 1936. Mental Tests Used in Genetic Studies: The Performances of Related Individuals in Tests Mainly Educative and Mainly Reproductive. Unpublished Master's Thesis, University of London.

Tran, A., Zeckhauser, R., 2012. Rank as an Inherent Incentive: Evidence from a Field Experiment. Journal of Public Economics 96 (9-10), 645–650.

Appendix

A Perception of the work tasks

In order to test if our experimental task manipulation works as expected, the subjects have to answer the following five questions in the final questionnaire. The answer options range from 0 (in the sense of "not at all") to 4 (in the sense of "very much").

- Question 1: "*How much did you enjoy working on the task?*"
- Question 2: "*How challenging did you perceive the task?*"
- Question 3: "*How much effort did you provide during the task?*"
- Question 4: "*How exhausting did you perceive the task?*"
- Question 5: "*Please indicate your desire to win.*"

For the first four questions we only analyse answers given by subjects who participate in the Control treatment as their perception of the task only depends on the task itself. In the Private and Public treatments, conversely, the answers could potentially be influenced by additional effects from the competitive environments. Question 5, however, can only be answered by subjects participating in the Private and Public treatments as only they perform in a competition and can thus exhibit a desire to win in their given task.

B Competitive preferences

The trait competitiveness measure by Brown et al. (1998) is a shortened version of Helmreich and Spence's (1978) trait competitiveness scale. The exact wording of Brown et al.'s (1998) 4-item scale is as follows:

1. I enjoy working in situations involving competition with others.
2. It is important to me to perform better than others on a task.
3. I feel that winning is important in both work and games.
4. I try harder when I am in competition with other people.

C Additional descriptive statistics and regressions

Table C.1: Performance in the Slider task: Correctly solved sliders across treatments

	Performance			Difference in performance	
Treatment	Ability phase	After kick-off	After revelation	After kick-off – Ability phase	After revelation – Ability phase
Control	14.180	15.840	16.080	1.660	1.900
	(4.448)	(4.372)	(4.125)	(2.730)	(3.183)
Private	12.422	15.800	16.044	3.378	3.622
	(5.092)	(2.997)	(5.187)	(3.863)	(6.184)
Public	12.425	14.825	15.100	2.400	2.675
	(4.018)	(4.877)	(5.856)	(4.556)	(5.484)

50 subjects participated in the Control, 45 in the Private and 40 in the Public treatments. The first three columns display mean values of performance. The mean values presented in the last two columns are calculated as the absolute difference between the (sub-)phases in question. Standard deviations are given in parentheses.

Table C.2: Determinants of the number of solved sliders

	After kick-off		After revelation	
	All treatments	Private vs. Public	All treatments	Private vs. Public
Competitive preferences	-0.030	0.152**	0.079	0.405***
	(0.086)	(0.072)	(0.076)	(0.098)
Competitive preferences x Private	0.181		0.363***	
	(0.112)		(0.127)	
Competitive preferences x Public	0.450***	0.269*	0.222	-0.273
	(0.144)	(0.136)	(0.159)	(0.167)
Ability	0.887***	0.444***	0.613***	0.221
	(0.107)	(0.113)	(0.119)	(0.145)
Ability x Private	-0.443***		-0.314*	
	(0.156)		(0.179)	
Ability x Public	-0.147	0.296*	0.020	0.223
	(0.150)	(0.154)	(0.227)	(0.257)
Winner after kick-off				0.588***
				(0.213)
Winner after kick-off x Public				0.374
				(0.369)
Expected to be winner after kick-off				0.075
				(0.213)
Private	0.253**		0.149	
	(0.122)		(0.160)	
Public	0.129	-0.123	0.087	-0.142
	(0.181)	(0.172)	(0.181)	(0.256)
Constant	-0.136	0.117	-0.090	-0.110
	(0.095)	(0.076)	(0.094)	(0.176)
Observations	135	85	135	85
R^2	0.50	0.39	0.34	0.34

All continuous independent variables and the respective dependent variables are standardised. Robust standard errors are reported in parentheses: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table C.3: Performance in the Picture task: Solved pictures across treatments

	Performance			Difference in performance	
Treatment	Ability phase	After kick-off	After revelation	After kick-off – Ability phase	After revelation – Ability phase
Control	12.449	10.939	8.714	-1.510	-3.735
	(2.517)	(2.164)	(2.160)	(2.190)	(2.280)
Private	11.840	11.700	9.820	-0.140	-2.020
	(2.216)	(2.013)	(2.371)	(2.080)	(2.360)
Public	11.267	11.689	9.200	0.422	-2.067
	(2.750)	(2.353)	(2.702)	(2.726)	(3.360)

49 subjects participated in the Control, 50 in the Private and 45 in the Public treatments. The first three columns display mean values of performance. The mean values presented in the last two columns are calculated as the absolute difference between the (sub-)phases in question. Standard deviations are given in parentheses.

Table C.4: Performance in the Picture task: Correctly solved pictures across treatments

	Performance			Difference in performance	
Treatment	Ability phase	After kick-off	After revelation	After kick-off – Ability phase	After revelation – Ability phase
Control	11.184	9.796	6.735	-1.388	-4.449
	(2.233)	(2.345)	(1.741)	(2.308)	(2.209)
Private	10.540	9.460	7.020	-1.080	-3.520
	(2.323)	(2.409)	(2.143)	(2.089)	(2.225)
Public	10.044	8.800	5.778	-1.244	-4.267
	(2.558)	(3.109)	(2.601)	(3.868)	(3.7983)

49 subjects participated in the Control, 50 in the Private and 45 in the Public treatments. The first three columns display mean values of performance. The mean values presented in the last two columns are calculated as the absolute difference between the (sub-)phases in question. Standard deviations are given in parentheses.

Table C.5: Determinants of the number of correctly solved pictures

	After kick-off		After revelation	
	All treatments	Private vs. Public	All treatments	Private vs. Public
Competitive preferences	-0.034	0.123	-0.123	0.106
	(0.125)	(0.129)	(0.127)	(0.111)
Competitive preferences x Private	0.190		0.254	
	(0.183)		(0.180)	
Competitive preferences x Public	-0.052	-0.223	0.197	-0.144
	(0.264)	(0.267)	(0.231)	(0.190)
Ability	0.532***	0.543***	0.272**	0.613***
	(0.132)	(0.136)	(0.121)	(0.138)
Ability x Private	-0.007		0.228	
	(0.186)		(0.184)	
Ability x Public	-0.425	-0.446	-0.355	-0.835***
	(0.294)	(0.286)	(0.249)	(0.231)
Male	-0.182	-0.130	0.066	0.192
	(0.161)	(0.210)	(0.167)	(0.203)
Overconfidence	0.044	0.064	0.042	0.203*
	(0.079)	(0.095)	(0.081)	(0.103)
Agreeableness	0.115	-0.072	-0.035	0.030
	(0.107)	(0.157)	(0.109)	(0.161)
Conscientiousness	-0.132	0.019	0.020	0.166
	(0.093)	(0.118)	(0.111)	(0.139)
Extraversion	-0.064	-0.162	0.071	-0.040
	(0.082)	(0.109)	(0.114)	(0.145)
Neuroticism	-0.006	0.074	0.171**	0.181*
	(0.075)	(0.087)	(0.082)	(0.092)
Openness	-0.106	-0.104	0.039	0.111
	(0.079)	(0.107)	(0.089)	(0.100)
Winner after kick-off				0.052
				(0.318)
Winner after kick-off x Public				1.066**
				(0.461)
Expected to be winner after kick-off				0.740***
				(0.225)
Private	-0.020		0.140	
	(0.166)		(0.165)	
Public	-0.201	-0.217	-0.457*	-1.017***
	(0.263)	(0.259)	(0.233)	(0.305)
Constant	0.125	0.089	0.033	-0.001
	(0.129)	(0.161)	(0.124)	(0.200)
Observations	144	95	144	95
R^2	0.24	0.24	0.22	0.43

All continuous independent variables and the respective dependent variables are standardised. Robust standard errors are reported in parentheses: $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table C.6: Determinants of the number of solved pictures

	After kick-off		After revelation	
	All treatments	Private vs. Public	All treatments	Private vs. Public
Competitive preferences	-0.000	0.100	-0.025	0.077
	(0.144)	(0.122)	(0.106)	(0.087)
Competitive preferences x Private	0.101		0.119	
	(0.189)		(0.134)	
Competitive preferences x Public	-0.205	-0.305	-0.004	-0.165
	(0.206)	(0.186)	(0.170)	(0.151)
Ability	0.508***	0.528***	0.328***	0.482***
	(0.157)	(0.098)	(0.098)	(0.115)
Ability x Private	0.020		0.099	
	(0.183)		(0.142)	
Ability x Public	-0.258	-0.278	-0.357*	-0.619***
	(0.239)	(0.202)	(0.194)	(0.208)
Winner after kick-off				-0.074
				(0.245)
Winner after kick-off x Public				0.882**
				(0.400)
Expected to be winner after kick-off				0.449*
				(0.230)
Private	0.329*		0.327**	
	(0.177)		(0.136)	
Public	0.288	-0.040	-0.103	-0.753***
	(0.197)	(0.181)	(0.211)	(0.268)
Constant	-0.222	0.107	-0.108	0.187
	(0.139)	(0.103)	(0.094)	(0.130)
Controlling for mistakes	Yes	Yes	Yes	Yes
Observations	144	95	144	95
R^2	0.30	0.31	0.41	0.44

All continuous independent variables and the respective dependent variables are standardised. Robust standard errors are reported in parentheses: $^{***}p < 0.01,^{**}p < 0.05,^{*}p < 0.1$

D Translated instructions for the Slider task

Instructions for the Experiment

[All treatments:]
Welcome to today's experiment. You receive 8 Euro for your participation.
It is of utmost importance that you do not talk to other participants during the course of the experiment. Furthermore, we would like to ask you to switch off your mobile phones.

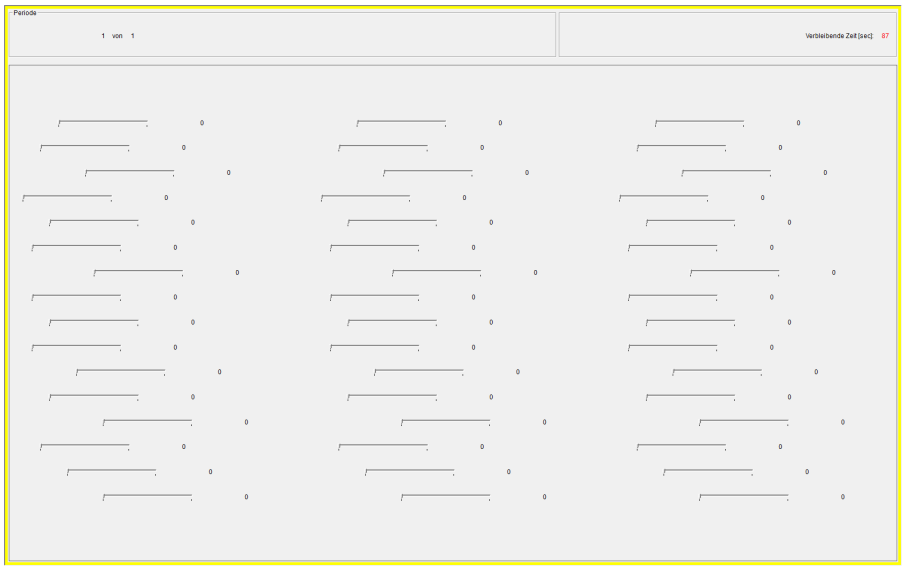
The experiment comprises two parts. Whenever a new part begins, your computer screen provides you with additional information about what is going to happen in the next part of the experiment. Please raise your hand, should you have a question concerning the instructions. We will immediately come to you to answer your question.
After the two experimental parts you will be given a short final questionnaire. Please answer all questions carefully.

The first part starts with a task in which you have to regulate sliders to position 50 using your computer mouse.
On your computer screen you will see many slider lines. At the beginning, each slider stands at position 0. By using the mouse they can be moved up to position 100.
You are given a total of 90 seconds to work on the task. During this period you can work up to a maximum of 48 sliders. Your task is to move as many sliders as possible to position 50. The task stops automatically after 90 seconds.

At the end of the experiment your computer screen informs you about your earnings from today's experiment.

On the next page you can find an example.

The following example illustrates your task. On your computer screen you are presented with several slider lines. The number right next to each slider line indicates the current position of the slider on the line. You can adjust a slider by clicking on it and shifting it with your mouse.



Your task is to adjust the sliders to position 50.

Instructions for Part 2

[All treatments:]
In the following part 2 you are going to work on the same task as in part 1.

[Control treatment:]
This part comprises two sub-parts. In each sub-part you are given 90 seconds to work on the task. During each of these periods you can work on a maximum of 48 sliders. Your task is again to move as many sliders as possible to position 50. The time for the first sub-part ends automatically after 90 seconds. A short break will follow before the second sub-part starts. In the second sub-part you work on the same task for further 90 seconds.

[Private and Public treatment:]
The only differences:

1. This part comprises two sub-parts. In each sub-part you are given 90 seconds to work on the task. During each of these periods you can work on a maximum of 48 sliders. Your task is again to move as many sliders as possible to position 50. The time for the first sub-part ends automatically after 90 seconds. A short break will follow, in which you receive new information before the second sub-part starts. In the second sub-part you work on the same task for further 90 seconds.

2. In this part of the experiment all participants are assigned to groups. Each group comprises 5 participants.

After finishing the two sub-parts a winner is determined within each of the 5-person groups. The person, who has completed the most patterns CORRECTLY over the two sub-parts is the winner in his or her group. All winners receive the information about being ranked as number 1 in their group via their computer screens. The remaining participants are only informed about not being ranked as number 1 in their group.

[In Public treatment additionally:]
The winners will be asked to come to the front of the lab so that all remaining participants can see them. All remaining participants applaud the winners. Furthermore, we take a photograph of every winner for a certificate. This certificate is going to be posted in the glass case in front of the laboratory in the last two weeks of July.
When you click on CONTINUE you will be presented with an example of how this certificate is going to look like.

[All treatments]
Please click on CONTINUE when you are ready to start working on the task.

E Translated instructions for the Picture task

Instructions for the Experiment

[All treatments:]
Welcome to today's experiment. You receive 8 Euro for your participation.
It is of utmost importance that you do not talk to other participants during the course of the experiment. Furthermore, we would like to ask you to switch off your mobile phones.

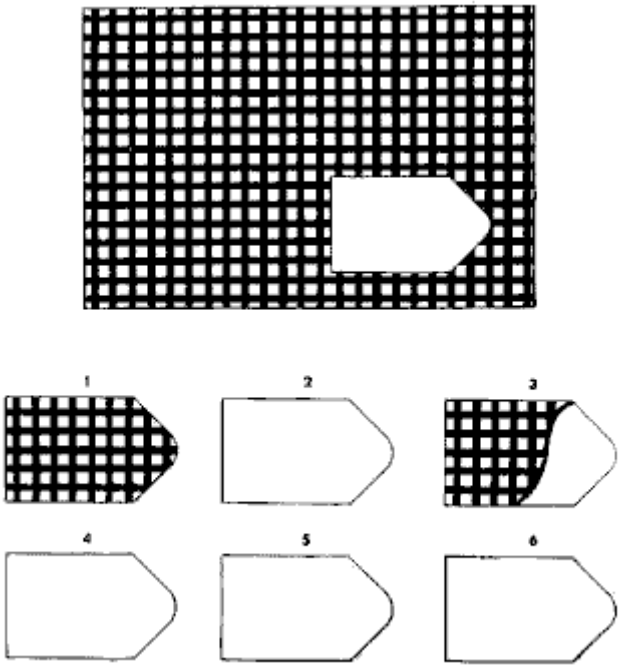
The experiment comprises two parts. Whenever a new part begins, your computer screen provides you with additional information about what is going to happen in the next part of the experiment. Please raise your hand, should you have a question concerning the instructions. We will immediately come to you to to answer your question.
After the two experimental parts you will be given a short final questionnaire. Please answer all questions carefully.

The first part starts with an excerpt of tasks that is used in intelligence tests to determine a person's intelligence quotient (IQ).
Your task is to complete patterns. You are given a set of potential elements from which you are supposed to choose the missing one. Whenever you have decided which element you want to choose, confirm your choice by clicking on the CONTINUE button. Subsequently you are shown a new pattern. Note that as soon as a new pattern is presented you can no longer change your answer to the previous one. In total you are given 90 seconds to work on the task. During this period you will be presented with a maximum of 18 patterns. Your task is to solve as many patterns as possible. The time ends automatically after 90 seconds.

At the end of the experiment your computer screen informs you about your earnings from today's experiment.

On the next page you can find an example.

The following example illustrates your task. On your computer screen you are presented a pattern. Below the pattern you can find the potentially missing elements.



Furthermore we ask you for the missing element on the same screen:

"Please choose the missing element for the pattern."

Next to the pattern you can find the answer options. Please tick the answer that you have chosen.

o 1
o 2
o 3
o 4
o 5
o 6

Instructions for Part 2

[All treatments:]
In the following part 2 you are going to work on the same task as in part 1.

[Control treatment:]
This part comprises two sub-parts. In each sub-part you are given 90 seconds to work on the task. During each of these periods you will be presented with a maximum of 19 patterns. Your task is again to solve as many patterns as possible. The time for the first sub-part ends automatically after 90 seconds. A short break follows before the second sub-part starts. In the second sub-part you work on the same task for further 90 seconds.

[Private and Public treatment:]
The only differences:

1. This part comprises two sub-parts. In each sub-part you are given 90 seconds to work on the task. During each of these periods you will be presented with a maximum of 19 patterns. Your task is again to solve as many patterns as possible. The time for the first sub-part ends automatically after 90 seconds. A short break follows in which you receive new information before the second sub-part starts. In the second sub-part you work on the same task for further 90 seconds.

2. In this part of the experiment all participants are assigned to groups. Each group comprises 5 participants.

After finishing the two sub-parts a winner is determined within each of the 5-person groups. The person who has completed the most patterns CORRECTLY over the two sub-parts is the winner in his or her group. All winners receive the information about being ranked as number 1 in their group via their computer screens. The remaining participants are only informed about not being ranked as number 1 in their group.

[In Public treatment additionally:]
The winners will be asked to come to the front of the lab so that all remaining participants can see them. All remaining participants applaud for the winners. Furthermore, we take a photograph of every winner for a certificate. This certificate is going to be posted in the glass case in front of the laboratory in the last two weeks of July.
When you click on CONTINUE you will be presented with an example of how this certificate is going to look like.

[All treatments]
Please click on CONTINUE when you are ready to start working on the task.

F Certificates used in the lab experiment

G Certificate used in the field experiment

Economics Working Papers

2013-15: Kaleb Girma Abreha, Valérie Smeets and Frédéric Warzynski: Coping with the Crisis: Recent Evolution in Danish Firms' - International Trade Involvement, 2000-2010

2013-16: Peter Arendorf Bache and Anders Laugesen: An Industry-Equilibrium Analysis of the LeChatelier Principle

2013-17: Anna Piil Damm and Christian Dustmann: Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?

2013-18: Christian Bjørnskov and Jacob Mchangama: Do Social Rights Affect Social Outcomes?

2013-19: Benoit Julien, John Kennes, and Moritz Ritter: Bidding for Clubs

2013-20: Ina Charlotte Jäkel: Import-push or Export-pull?

2013-21: Tor Eriksson, Jay Pan and Xuezheng Qin : The Intergenerational Inequality of Health in China

2013-22: Martin Paldam: How do partly omitted control variables influence the averages used in meta-analysis in economics?

2013-23: Ritwik Banerjee: An Evaluation of the Revenue side as a source of fiscal consolidation in high debt economies

2013-24: Nabanita Datta Gupta, Mona Larsen and Lars Brink Thomsen: Do wage subsidies for disabled workers result in deadweight loss? – evidence from the Danish Flexjob scheme

2013-25: Valerie Smeets, Sharon Traiberman and Frederic Warzynski: Offshoring and Patterns of Quality Growth: Evidence from Danish Apparel

2013-26: Torben M. Andersen, Jonas Maibom, Michael Svarer and Allan Sørensen: Do Business Cycles Have Long-Term Impact for Particular Cohorts?

2013-27: Martin Paldam: Simulating publication bias

2013-28: Torben M. Andersen and Allan Sørensen: Product market integration, tax distortions and public sector size

2014-01: Leonie Gerhards and Neele Siemer: Private versus Public Feedback - The Incentive Effects of Symbolic Awards