I. Introduction

Concerns about false incentives arising from performance payments are more relevant than ever in these current times of public discussion about managerial excesses and corruption. The financial crisis seems to have dealt a hard blow to the acceptability of bonuses, piece rates, and shares as forms of compensation. Yet economists have been arguing for a long time in favor of performance payments in contrast to purely fixed wages. However, the economic literature is far from being unambiguous about the incentive effect of variable payments. On the one hand, for instance, Lazear (2000) provided evidence for an increase in performance due to incentive and selection effects through the implementing of performance payments. On the other hand, some analyses even find a reduction of work effort if a monetary incentive is attached (e.g. Gneezy/Rustichini 2000, Pokorny 2008). The so-called crowding-out effect is based on the idea that an extrinsic incentive, such as a piece rate, can extinguish intrinsic motivation (Frey 1997). Furthermore, concerns about the long-term effect of variable compensation exist because its time horizon is usually much shorter than that of the goals of firm owners (Agrawal/Knoeber 1996). But are there any alternatives for aligning employers' and employees' goals and reducing the agency problem?

Regarding a survey by Egon Zehnder International in 2009, managers judge, for instance, the allocating of responsibility and the provision of feedback to be among the top criteria for work motivation, even more so than the salary itself. Non-monetary incentives (for an overview, see Lazear 1998, Ellingsen/Johannesson 2007) might be superior to money because they show recognition and attention-giving by supervisors and make employees feel useful and appreciated which could lead to affective reactions (Shaffer/Arkes 2009). A common way to express recognition is via the implementation of simple distinctions, such as 'Employee of the month'. Beyond being a friendly gesture of the supervisors, this kind of public feedback can also create a raise in productivity due to peer (Falk/Ichino 2006, Mohnen et al. 2008) and signaling (Frey/Neckermann 2010) effects. Furthermore, a symbolic reward might not reduce but enhance intrinsic motivation because winners and losers can rate the award differently due to a missing objective value (price). In addition it seems to be an inexpensive way to create tournaments where employees strive for the honor of being 'the best'. The idea of status-consciousness and the advantage of immaterial incentives being completely sharable between agents and principals have been discussed at length in the sociology literature (Sorauren 2000, Ball/Eckel 1998).

However, only recently have economists joined the debate on analyzing whether symbolic rewards really do work and under which circumstances. Empirical studies (e.g. Neckermann et al. 2009, Frey/Neckermann 2008 and 2009) show the high prevalence of awards and provide evidence for a positive impact on performance. Analyzing panel data from the call center of a large international bank, Neckermann et al. (2009) find an increase in the performance of award recipients compared to non-recipients, although the award is not directly linked to performance itself but given for non-contractible work, such as volunteering to work during vacation periods. They identify publicity, deliberately vague evaluation criteria, unenforceability, and a tournament character as essential for successful incentive modeling. In a vignette study, conducted at IBM, Neckermann and Frey (2008) were able to show that the probability of contributing to a public good at one's own expense significantly increases if the prospect of an award with an additional cash payment and high publicity is given. The authors argue that an additional cash payment attached to an award shows the significance and seriousness of the given reward. The empirical study of Blanes i Vidal and

Nossol (2009), in contrast, provides evidence for a motivational effect of tournaments without prizes or even a symbolic award. They argue that information about how employees perform relative to their colleagues is an incentive in itself. Concerns about relative standing, even if this information has been privately disclosed, seems to be followed by a striving to be the best or not the worst worker on staff. Relative standing is also important for incentive effects of wages, which has been shown, for instance, by Clark, Masclet, and Villeval (2010). Blanes i Vidal and Nossol (2009) differentiate between two incentive effects of performance evaluation. The first is called the 'kickoff moment'. This is the point when the evaluation period of employees starts. The second one is called the 'revelation moment'. At this point, the first evaluation period has ended and ranking positions based on their work performance are privately disclosed. These findings show that it is not quite clear to what extent an award has to be valuable in monetary terms and how a certain degree of publicity may drive the results.

Models which try to give reason to empirical findings on the influences of non-monetary incentives by adapting theoretical approaches have been developed by, for example, Ellingsen/Johannesson (2008), Dur (2009), Moldovanu et al. (2007) or Auriol/Renault (2008). Ellingsen and Johannesson implement self-esteem in agents' utility functions, which depends on the type of principal. If a principal can signal that she or he is the kind of person who is worth impressing, agents benefit more from a principal´s high degree of recognition than they would do if they did not respect him or her themselves. Dur´s (2009) argumentation is quite similar. He defines 'attention' as a substitute for high payment, and constructs a model where principals can signal their true solicitousness for employees by offering a lower wage than that offered by selfish principals in a separating equilibrium. Auriol and Renault (2008) instead advise against the understanding that recognition is an inexpensive substitute for money. The success of recognition depends on its sincerity and credibility. These properties can only be assured when recognition is earned honestly and is sufficiently hard to achieve. The authors stress as well that status is a scarce resource that depends on the hierarchies of organizations. The tournament character of awards also implies that there are not only winners but also losers. Therefore, the positive motivational effect of differentiating between employees might be positive for those employees with a higher status but could also reduce motivation for those for whom rewards are out of reach (Eriksson et al. 2009).

In this paper, we try to shed light on the effect of a non-monetary and non-material incentive and its effect on work performance. Our focus is on the impact of status revelation induced by a symbolic reward. According to Ball et al. (2001), we define a person´s status as *"a ranking in a hierarchy that is socially recognized and typically carries with it the expectation of entitlement to certain resources"*. The resource of our experiment is not valuable on its own but might be desirable because of its trophy value. This means that even if an award is worthless in monetary regards, it is a lasting reminder of status to oneself and to others. Ball et al. (2001) were able to find in their experiment that in a market setting with buyers and sellers, the side of the market that has a higher status than its counterpart captures a greater share of the surplus.

Our design, in contrast, is without interaction and has an anonymous and abstract setting. We test the effect of an award for solving simple mathematical equations. The awards are designed as buttons in the colors gold, silver, and bronze. All participants work on their own and no money is at stake. We compare two treatments under fixed- wage conditions, hereafter called 'control treatment' and 'award treatment'. The control treatment, on the one hand, is a working situation where performance evaluations and relative standings are only privately disclosed. On the other hand, the award treatment ends with a public award ceremony disclosing their relative standing to all participants. To our knowledge, analyses of private and

public performance evaluations under comparable settings with a focus on status revelation of symbolic awards, do not exist up to now. The 'award effect' could often not be distinguished from the 'feedback effect' in empirical studies because the two go hand in hand. Our design instead allows an isolated examination of each effect. Furthermore, the lab experiment enables us to control for other factors which might have an influence on existing empirical results summarized in the section above. Hence, we control for, for example, ability or personal characteristics in our experiment.

Our results show that implementing an award does influence effort, but that this effect is not positive under all circumstances. We were able to single out the effect of group size in tournaments and the relative standing in a working group as relevant factors for motivation. The positive effect of awards is that they enhance the motivation of employees who are already top performers, but do not have this motivational effect on bad ones, owing to discouragement. In small groups, awards have a higher incentive effect because status is more exposed and the rewards are more exclusive than in larger groups.

The structure of our paper is the following. In the next section, we will continue with the deduction of our hypotheses and the presentation of our experimental design. Afterwards we will point out the key results for all subjects and follow them up with analyses for the different performance groups. We finish with a conclusion and a forecast for further questions of research.

II. Design, Method, and Procedure

2.1 Hypotheses and Experimental Design

Our experiment is divided into three rounds. In each round, subjects had the choice of solving simple mathematical equations or reading interesting articles (see Figure 1). We provided articles (ten topics from politics, sports to yellow press) to ensure that people had an outside option and did not choose to work simply because no alternative was given. We measured performance by the number of correctly solved simple equations, and we differentiated between both the quantity (number of worked equations) and the quality (number of errors) of work. We chose mathematical equations because people in general associate mathematical skills with intelligence and therefore strive to be one of the best. We also asked our subjects' opinions of the given task being a good predictor of intelligence, and 29% agreed that it was. The agreement, though, is significantly correlated to the achieved ranking position of each round. The experiment started with a screen on which all subjects could check their solutions to the five simple test equations. This was to ensure a basic understanding of our task, and in addition we were able to use the number of wrong answers as an ability check. The result of this test, however, had no effect on payment. After the first round, the subjects were asked to give a self-assessment of their performance in the first round by choosing their ranking position (the best 5%, 30%, 45%, or the worst 20%). This discrete classification is due to the indivisibility of tangible awards in our award treatment and is based on the ranking used by Jeffrey (2009). The self-assessment allowed us to control for overconfidence or underestimation. We rewarded a correct assumption with 2 Euros[1] to give an incentive for subjects to make their best guess.

[1] 1 Euro = 1.27 US-Dollar

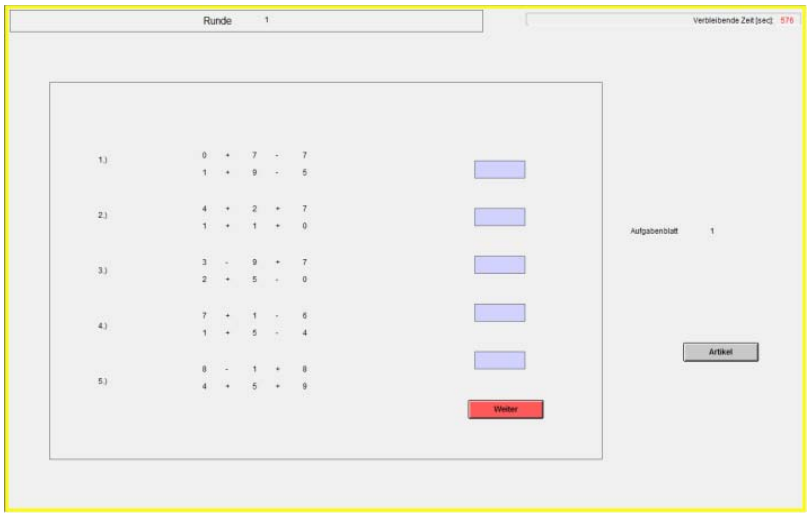

Figure 1: Exemplary screen of the mathematical task used in the experiment

To test which differences in incentives occur according to a symbolic reward, we ran two treatments which were completely equal in round one but were differentiated afterwards. The equal design in the first round was necessary for an additional measure to control for differences in abilities between and within our two subsamples. In the *control treatment*, all rounds were identical except that the subjects got their results after the second and third rounds (number of worked equations, the share of correct solutions and their ranking position (same classification as before)). In our *award treatment*, we told the participants that their ranking position would be publically disclosed to everybody after the experiment in a ceremony, where they would be awarded buttons in bronze, silver, or gold according to their relative positions (for an illustration, see Appendix). The worst 20% would receive no button. To keep the incentives equal in the second and third rounds, we told subjects that the ranking position which mattered for the ceremony would be decided by the flip of a coin.

Each round lasted for 10 minutes, which led to a certain degree of tiredness and therefore considerable costs of effort. However, we designed the third round not only for modeling some kind of long-term effect but also for testing the motivational effect after the real ranking position had been disclosed to the participants. In both treatments, only a fixed wage was paid, without any variable pay attached to the work performance. The only additional payment was the belief incentive of 2 Euros. Thus, our first hypothesis, according to the standard theory, is that the minimum effort level should be chosen in both treatments under fixed-wage conditions.

H1: No monetary incentive exists according to standard economics. Hence, the level of effort in both treatments should be low and there should be no differences between treatments.

As explained before, the purely symbolic award given in our second treatment might induce an implicit trophy value by reason of status disclosure. Based on feelings of pride or shame and a concern about what others might think, we assume a higher work effort in our award treatment (Ellingsen/Johannesson 2008). The chosen awards were buttons in the colors of Olympic medals and therefore their meaning should be clearly depicted by the linkage to sport tournaments. In addition, the award was like a certificate or a clap on the shoulder:

meaningless in monetary terms and therefore only valuable in a symbolic way (Kosfeld/Neckermann 2010).

H2: A symbolic award leads to higher work performance due to feelings of pride or shame through status disclosure.

Following Auriol and Renaults (2008), we do not expect this motivational effect to be equal over all subjects. For one thing we follow their argumentation of status disclosure to increase work motivation for top performers, who have a reasonable chance of being winners of this tournament, whereas bad performers might even be discouraged by the prospect of losing a tournament. Therefore, our prediction is that the motivational effect is not unambiguous but depends on the relative standing that a person already holds. Following Blanes i Vidal and Nossols (2009), we also differentiate between the kickoff moment (beginning of round two) and the revelation moment (beginning of round three).

Kickoff moment:

H3: Some intrinsic motivation could be driven by the fact that all participants get their results in the second and third rounds. We expect this motivation not to be extinguished but enhanced by the prospect of an award for top performers, who have a reasonable chance of being one of the winners of the tournament.

Revelation moment:

H4: After the results of the second round have been disclosed, good performers should feel encouraged in their work effort, whereas bad ones might instead be discouraged. This lack of motivation for bad performers should be diminished by an award because of the fear of public disgrace.

In addition to the experiment itself, we let the participants fill in a document with personal data (e.g. gender, nationality, age). In this way, we controlled for influences other than the given incentives by awards on effort, and created variables to control for these effects in our regressions. Besides, we had ex post questions for investigating the motive of effort decisions.

We used the program z-tree by Fischbacher (2007) and recruited our subjects with e-mails sent to students via our online learning rooms, and ads posted all over the campus. Students from various departments participated and in total we had 80 subjects over three sessions. The average age was 26 and 64% were male (see Table 1 in the Appendix). A chi-square test on the number of male participants and the mean age in each treatment showed no significant difference by Fisher's exact test ($p=0.486$ and $p=0.309$)[2]. The payment at the end of the experiment was 10 Euros per hour and the experiment took 1 hour and 30 minutes. The experiment took place in the lab for experimental studies at RWTH Aachen University between November 2009 and February 2010.

[2] The results are robust for Pearson's chi-square test.

2.2 Real Effort Task

As we have mentioned before, the real effort task allows us to differentiate between quantity and quality of work, and seems to be sufficiently accepted as a performance task to justify a tournament for being 'the best'. As can be seen from Figure 2, the number of worked and solved equations (scores) rises over time. So by just looking at means, the motivational and learning effects seem to outshine tiring effects, as the number of worked equations rises from 45 to 54. But if we take a closer look at the percentage of correct answers (Table 1), a slight effect of diminishing concentration can be found, as the percentage decreases from 90.19% in the first round to 83.16% in round three. The decrease in work quality could also, of course, be a result of stress due to the incentive of result disclosure in rounds two and three. However, the award implementation does not have a significant negative effect on work quality according to a Wilcoxon rank-sum test. The assumption that subjects feel tired by round three is thereby supported by looking at the usage of article reading as an outside option in Table 1. The mean time spent on article reading drops in the second round from 1.63 minutes to 1.20 minutes due to the kickoff moment, but increases again to the original size in round three.

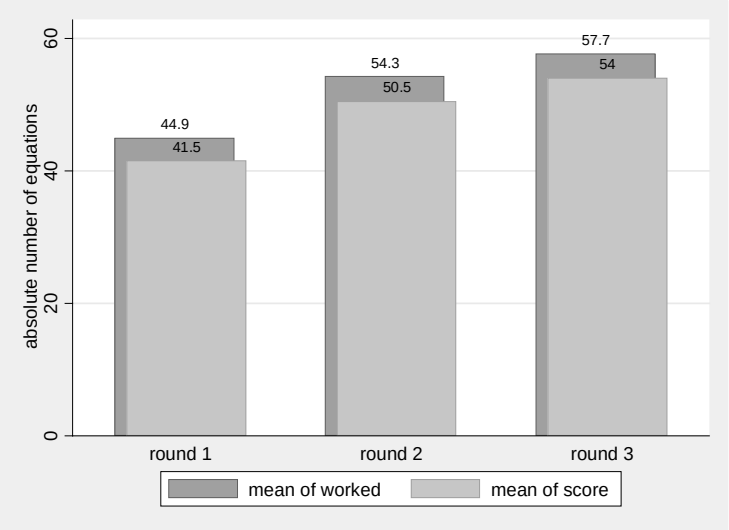

Figure 2: Means of worked equations and correct solutions in each round [3]

2.3 Ability Check and Gender Differences

We are aware that ability has a certain impact on the effort task, even if it requires only basic mathematical skills. Therefore, we have three different measures of ability. The first one is the sum of all errors and the second is the time it takes to solve all five test equations. No subjects were permitted to leave the test stage until they had got all equations right. The time ranges from half a minute to over ten minutes. The high deviation is reflected in the number of errors as well. An even better measure of ability is the score number in the first round. As can be seen in Figure A1 in the Appendix, the score density in round one is normal distributed, with its mean at 42 solved equations. The standard deviation for the score density of the second

[3] The subject who filled in over 200 simple equations in the second round is dropped. In the third round he returned to an earnest solving strategy.

and third rounds is higher and the number of subjects who do not work the equations increases.

Regarding ability differences between gender, we ran a Wilcoxon rank-sum and t-tests over both treatments and could not find significant differences concerning scores in round one but did find a significant higher test time for women in the Wilcoxon rank-sum test (p=0.021, t=0.17). This holds also if we exclude the outlier of ten minutes test time, who happened to be a woman. Considering gender differences over all rounds, we find interesting results. As stated before, scores are not significantly different between gender (p=0.125, t=0.281) but women seem to be more cautious concerning work quality, as expressed by a significantly higher percentage rate of correct answers (p=0.011, t=0.018), (Figure A2). This is robust if the male participant who simply filled out over 200 equations with the same number is excluded. These results are in line with previous findings that men feel a higher pressure to be among the best and as a result neglect quality in favor of higher quantity (Delfgaauw et al. 2009). Nevertheless, women's usage of the outside option is significantly lower (p=0.015, t=0.015) and declines with each further round (Figure A3) so that it could not be said that women avoid being part of a tournament (as has been found in Niederle/Vesterlund (2007)). Both Figures are included in the Appendix.

III. Results

3.1 Non-Parametrical Tests

Now we want to take a closer look at the incentive effects of awards. As can be seen in Figure 3 the subject pool in our award treatment starts with a lower score in round one but the increases from each round to the next one are slightly higher so that these subjects result in being quite as good as the subjects in our control treatment. This gives us a first indication that the motivational effects might differ between both settings. Nevertheless we do not find a significant difference by using a Wilcoxon rank-sum test. Furthermore, no significant differences between the numbers of wrong answers exist. However, as other experimental evidence has shown before, we can reject the first hypothesis following the standard theory that nobody would work under fixed wage settings.

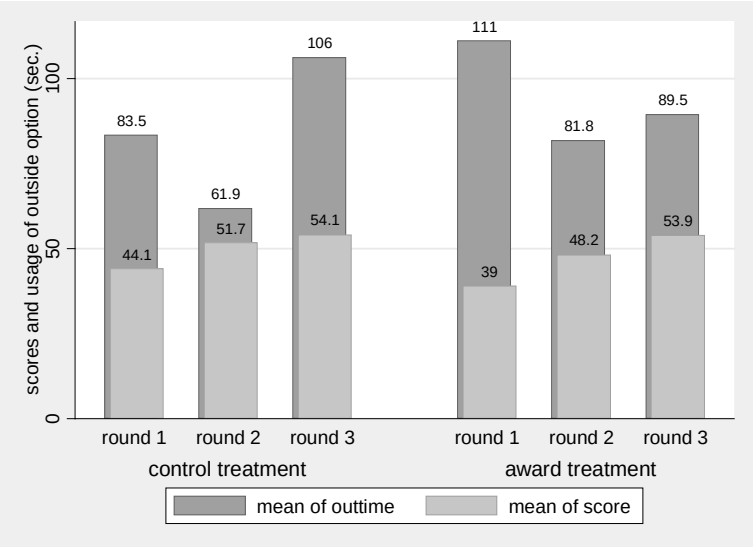

Figure 3: Treatment differences in time of article reading and number of correct solutions

Puzzled by the results in our pretest in November 2009, which had confirmed our second hypothesis, and the divergent results followed by our session in December 2009, we decided to take a closer look at the aspect of group size on the motivational effect of awards. We suspect our results were driven by a smaller group size in our pretest (consisting of 9 and 7 subjects in our award and control treatment, respectively) in contrast to the group size of 22 participants in our main session in December. Therefore, we conducted one last session in February 2010 to test for this effect by reducing the size of our sessions to 10 participants per treatment. The results differentiated by group size and treatments are shown in Figure 4.

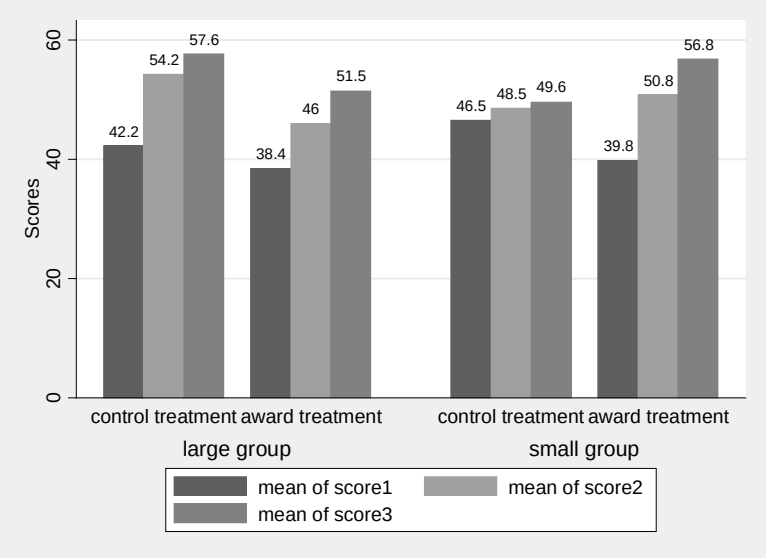

Figure 4: Mean scores in large and small groups

Whereas the increases in work performance (measured in solved tasks) are high in both treatments, the motivational effect is significantly larger when implementing an award in our small-group sessions. This effect is significant for performance differences between rounds one and two at the 10% level (p= 0.078) based on a Wilcoxon rank-sum test. The overall increase of performance between the first and third rounds is nearly significant (p=0.109). Aligned with findings in small groups, the time of article reading decreases from the first to the second round in the award treatment, whereas in the control treatment it slightly increases. This difference is significant at the 5% level (p=0.045). Looking at article reading in the large group in Figure 5, there is a sharp decrease of outside option usage in the second round in the control treatment, and afterwards an increase which leads to the mean time of the first round. In contrast, there is a slight decrease of article reading in the award treatment. Overall, the increase in time-spending on article reading from the second to the third round is higher in the control treatment, even if this does have to be interpreted with caution as it is not statistically significant.

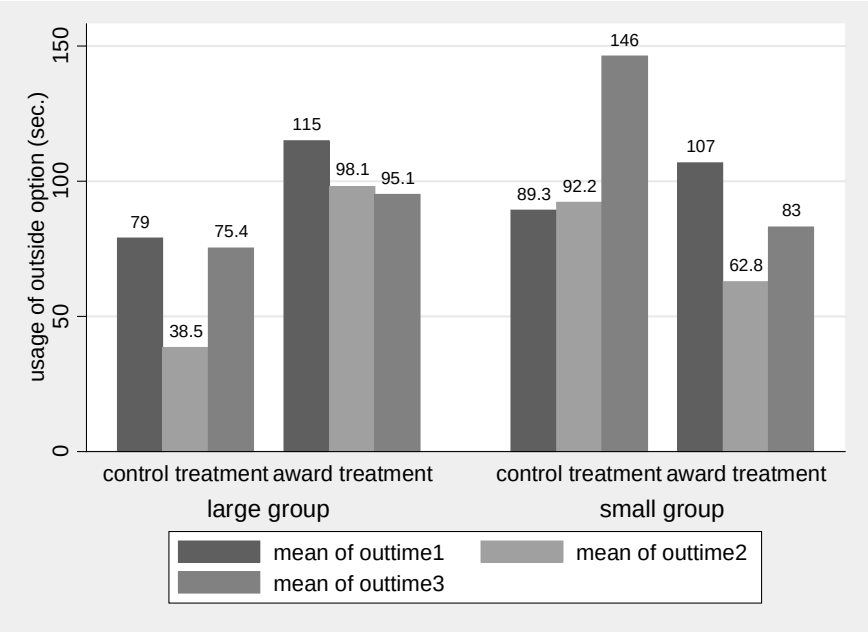

Figure 5: Mean time of article reading in large and small groups

3.2 Multivariate Analysis

Hence, we run multivariate OLS regressions to check whether the incentive effect of awards is robust when personal characteristics of our subjects, such as gender, age, and ability are included. As a measurement of work performance, we use on the one hand the absolute scores (number of correct solutions), and on the other hand the differences between the two rounds as dependent variables. Table 2 contains the regression results on the absolute values of scores. The first three Models show the effects on the second and the last three Models on the third score. In all Models we find a highly significant effect of ability measured with the first-round score. The other two ability measurements are insignificant, except for a negative effect of the test time on the third score. The award dummy shows that awards in general do not have a significant effect on scores and even show a negative sign. We also include a dummy for sessions with small numbers of participants (smallgroup) and an interaction term (smallaward) to test the incentive effect of awards in these sessions. The coefficient of the interaction term is significant and positive and shows a high magnitude. It has to be interpreted as an increase in work performance by 12 additionally solved equations (Model 3) in round two and 16 additionally solved equations (Model 6) in round three by means of a given award in small-group settings. The third and sixth Models also imply two dummies for a question about subjects' motivation, included in the questionnaire at the end of every experiment. The question read as follows: "Which of the following was essential for enhancing your work input?" Several answers were given (such as 'having fun' or 'feeling some kind of exam pressure to solve math tasks') and subjects could choose more than one. The two incentives which show a high impact on work performance are the incentives to be one of the best and not to be one of the worst performers. Both motivations have a significantly positive effect on work effort even if this effect is stronger for subjects striving to be one of the best ($p=0.001$).

Table 2: OLS regressions with scores as dependent variables

Dependent variable	**Score 2**			**Score 3**		
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
score1	0.653***	0.692***	0.497***	0.799***	0.732***	0.499**
	(3.93)	(3.71)	(3.13)	(4.05)	(3.08)	(2.36)
test time	-0.055	-0.042	-0.036	-0.06	-0.069*	-0.068*
	(-1.57)	(-1.16)	(-1.19)	(-1.59)	(-1.68)	(-1.9)
test error	-0.065	-0.153	-0.052	0.155	0.089	0.231
	(-0.21)	(-0.48)	(-0.21)	(-0.77)	(-0.37)	(-0.93)
award	-7.748	-7.942	-5.492	-5.023	-5.675	-4.644
	(-1.42)	(-1.4)	(-0.84)	(-0.67)	(-0.74)	(-0.54)
smallgroup	-9.419	-9.721	-7.868	-13.009	-11.887	-10.061
	(-1.52)	(-1.62)	(-1.59)	(-1.66)	(-1.53)	(-1.49)
smallaward	14.522*	14.705*	12.177*	18.274*	18.392*	15.8*
	(-1.68)	(-1.72)	(-1.72)	(-1.72)	(-1.7)	(-1.73)
age		-0.172	0.263		-0.299	0.18
		(-0.74)	(-0.96)		(-1)	(-0.5)
male		5.59	6.653*		-1.792	-0.517
		(-1.3)	(-1.75)		(-0.39)	(-0.11)
overconfidence		2.546	-0.87		-4.393	-8.005
		(-0.54)	(-0.22)		(-0.64)	(-1.32)
awbest			0.654			3.9
			(-0.08)			(-0.39)
being the best			22.105***			23.676***
			(3.64)			(2.84)
being not the worst			7.82*			11.187**
			(-1.91)			(2.22)
Observations	80	80	80	80	80	80
R-squared	0.44	0.46	0.64	0.39	0.4	0.57

Robust t-statistics in parentheses; significant at 10% (*), 5% (**) and 1% level (***). Dependent variables are the absolute score values (number of correct solutions) in round two and three. Constants are included.

Following the experimental design of Jeffrey (2009), and to verify whether our results are robust, Table 3 (in the Appendix) shows the OLS regression with differences in scores as dependent variables. The 240 observations over three rounds are clustered for the 80 subjects. By using the score differences we control for latent skill differences between subjects and therefore reduce unexplained variance. The first two Models confirm the results of Table 2 except that the dummy for small group sessions is significantly negative. It seems that the readiness to increase work effort was smaller in these groups but this effect is overcompensated by the high increase of work performance in the award treatment. In addition both dummies for motivational intentions and being male have a positive effect on the differences between both rounds. Looking at the differences in scores of the second and third round we find a significantly lower increase (or even decrease) for subjects who needed longer at the test stage. Furthermore Models three and four show two interesting findings. First, men seem to get bored or tired between the second and third rounds and reduce their work effort. Second, subjects who overestimated their ranking position after the first round reduce their work effort as well. By including an interaction term of overconfident men, both gender and overconfidence become insignificant. However, the interaction term is significant

at the 10% level (p=0.061) with a coefficient of -11.25 (table not included). This means that men rather than women seem to be more discouraged by the disclosed status in round two if they had been expecting to be better than they actually were. The fifth and sixth Models show the differences in scores from rounds one to three. Here, the effects are rather similar to the ones of the first two Models. In contrast to our intuition, test errors have a positive effect on work increase when motivational parameters are included. The motivational effect of awards seems to hold even in the long run for small groups. Hence, the interaction term between striving to be one of the best and the award treatment, awbest, is not significant in any Model. Therefore there seems to be no additional incentive for ambitious subjects to increase effort if an award is given.

So, considering our second hypothesis, it can be said that rewards, even if only of symbolical nature, do have a positive impact on work performance but that this effect should not be generalized. In particular, the design of tournaments and the breadth of hierarchies seem to play crucial roles in incentive settings. Awards do not work if they are not exclusive enough or if the lower levels of hierarchies are so broad that subjects do not feel exposed but can justify a bad performance with the argument of their being one among many. These effects of awards are shown in differences in sign and significance between small- and large-group sessions in our experiments. The small groups are less anonymous due to 'narrower' hierarchy levels. Instead of 1, 7, 10, and 5 subjects in the first, second, third, and fourth ranking groups, there were only 1, 2 or 3 respectively, 4, and 2 subjects on hierarchy levels in the small-group sessions. Thus the absolute number of good and bad performers is much smaller and this goes hand in hand with a stronger feeling of distinction or disgrace connected to the ranking group and enhanced by the publicity of an award.

3.3 Performance Groups

To analyze the difference in incentive effects between the four performance groups (the best 5%, 30%, 45%, or the worst 20%) we look at the mean scores of rounds one to three and the improvements in work performance between the three rounds in Table 4. First of all, we see a large variance of work performance in the first round from 14.5 to 67 solved equations in ten minutes. The differences between two rounds, however, do not differ significantly between the four performance groups by a Kruskal Wallis test. Even if it is not possible to observe cost functions in a real effort experiment, this analysis provides evidence that the improvements were not easier or harder for participants of verifying latent abilities on the task and confirms the reliability of differences in scores as dependent variables of Table 3 (see Appendix).

Table 4: Mean scores and performance improvements of the four performance groups

Performance groups	Score1	Improvement from round 1 to 2	Score2	Improvement from round 2 to 3	Score3
5% best	67.333	7.167	85.833	9.333	97.83
30%	56.000	2.333	68.500	7.042	76.041
45%	38.743	10.857	46.029	0.800	49.111
20% worst	14.467	12.933	14.933	5.000	10.143

The first two columns are clustered by the first-round performance group. Columns three and four are clustered by the second-round performance group and the last one by groups of round three. All scores are significantly different between groups (1% level) in a Kruskal Wallis test, whereas the improvements in work performance are not significantly different between performance groups (p=0.307; p=0.169).

Due to small observation numbers in the first and last ranking groups (only 6 observations in the former) we cluster the first and last two groups to best 35% and worst 65%. We replicate the OLS regression in Table 1 differentiated by the two performance groups. The results are shown in Table 5 (see Appendix). To start with, we look at the first two columns with score 2 as dependent variables. Following Blanes i Vidal and Nossols (2009), we have defined the instruction screen before the second round as the kickoff moment because the status disclosure after this round, privately or through an award, is implemented. We find that ability is, for good performers, significantly influential in all measurements. Furthermore, these performers are significantly better in large groups with a positive effect if they wanted to be one of the best. For worse performers, the motivation of wanting to be one of the best is accompanied by a fear of being one of the worst performers. The interaction term for awards in small groups is only significant for good performers. This result supports our third hypothesis that only good performers strive to get the award by increasing their work performance due to their chance to show their high status in relation to the other employees. As can be seen in Figure 6, top performers even reduce their work effort in the second round compared to the first one in small groups. In this regard, the higher increase in work effort for large groups in the control treatment is odd. One explanation for this might be the incentive effect of privately disclosing the results and relative standings at the end of this round. Because no other incentives in monetary or non-monetary forms were attached to the performance groups, we see our results in line with the evidence given by Blanes i Vidal and Nossols (2009). They found that the prospect of performance evaluation in the form of relative standings, such as tournaments without incentives, do work on their own. And it seems that an external incentive, such as a symbolic award, might even diminish this effect due to the crowding out of intrinsic motivation. Therefore, we find positive evidence for our third hypothesis, if only for small groups.

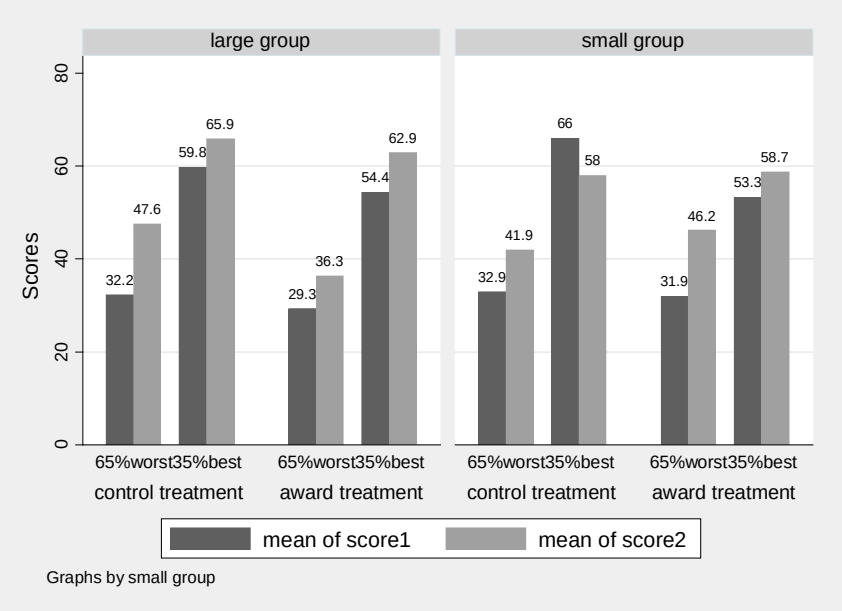

Figure 6: Scores before and after the kickoff moment for top and bad performers clustered by score rankings in round one

By comparing the differences in scores between the two performance clusters, we find a significantly larger improvement in scores for top performers than for bad performers from the second to the third rounds (see Table 4). This effect, in contrast to the improvement between the first two rounds, is significant at the 5% level by a Wilcoxon rank-sum test

(p=0.04). Hence, the revelation moment leads, as expected in hypothesis four, to higher work enhancements of good performers. In Figure 7, one can see that awards seem to diminish the discouraging effects of bad performers in the third round, and even lead to an increase in performance compared to the control treatment. But taking a look at the influences of the third score in Table 5 (see Appendix); we find no significant influence of the symbolic award at all. This finding is based upon an overrepresentation of third-group performers (70%) compared to fourth-group performers (30%) in the 65% worst cluster. Whereas subjects in the fourth group increase their performance by 5 additional equations on average, subjects in group three show only a work increase of 0.8 equations (see Table 4). Therefore, it can be assumed that being in the third group is quite acceptable to those subjects and they seem aware of their great distance away from the top but also seem to be sure to be good enough not to drop to the 20% worst performers. These subjects, instead, show some effort to avoid being tail enders. Unfortunately a differentiated analysis of all four categories is not possible due to our sample size. The dummy for overconfidence is significant for the participants who are top performers in the second group (Table 5). These participants work significantly less in the third round. But this has to be interpreted with caution because this change of performance group is only true for six participants (three in each treatment). Over both performance groups, the score in round three is nearly exclusively determined by the score in round two. All effects stay more or less the same if we divide the performance groups not into the ranking groups but into above- and below-average performers. Interestingly, the gender dummy becomes significant for the below-average performer. So as we have seen before, male participants are working less in the long-run in the absence of incentives. But this finding only holds for bad performers. Good performers, male or female, have a lasting interest to be one of the best. In conclusion, we could only support parts of our fourth hypothesis. Over all treatments, there seems to be a lower performance increase for bad performers compared to that of the good ones. However, we could not find a significant diminishing effect on discouragement by the implementation of an award.

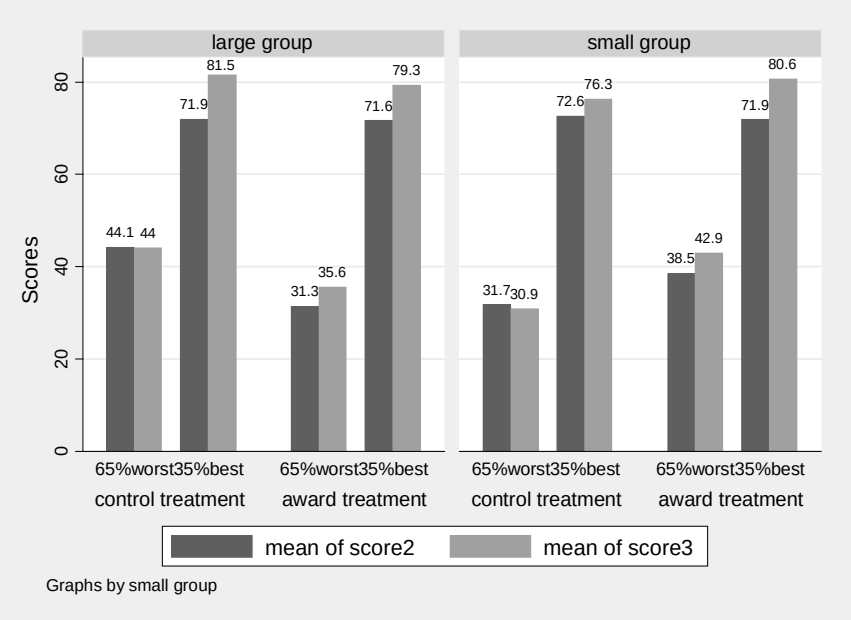

Figure 7: Scores before and after the revelation moment for top and bad performers clustered by score rankings in round two

IV. Conclusions

In this paper, we have examined the effect of a symbolic reward on work performance. We used the controlled setting of a laboratory experiment to exclude other effects, such as career concerns or reciprocity, which occur in work settings and might bias empirical investigations. We are aware that due to the anonymity of participants and the shortness of the work period, conducted by three rounds of ten minutes each, our design is abstract and not directly transferable for practical impacts. Nevertheless, we presume that in an ongoing work relationship with more social interaction, the values of awards would be even higher. Especially if employees feel bond with their supervisor and have the feeling that the company appreciates them, performance is of higher value, as has been shown by Ellingsen and Johannesson (2008). The experimental outcomes are therefore strong evidence of the incentive effect of purely symbolic awards due to an implicit trophy value. However, these findings are limited to our small-group settings, so that we conclude that the specific design of tournaments and constellation of hierarchy levels is essential for positive incentives and, if conducted incorrectly, such tournaments could even have negative implications. In contrast to standard theory, concerning the motivational effects of fixed wages, we observe a high work effort in general across treatments and a positive effect of the kickoff moment after the first round. So even if results are privately disclosed, people tend to work more. This effect is enhanced for good performers by the public disclosure of an award. Moreover, after the results have been disclosed, good performers also increase their performance in the third round significantly more compared to performers of the last two ranking groups. This effect, however, does not differ between both treatments.

What implication could, then, be extracted from these results? Because of recent incidents in the financial sector, monetary bonuses have become more and more discredited. However, economists insist on bonuses and piece rates being necessary forms of compensation to implement work incentives. With this paper we want to contribute to research which is focused on alternatives to monetary incentives. Awards are a cost-saving way to incentivize a workforce and, if conducted correctly, do not extinguish intrinsic motivation. But due to its tournament character it is not positive for all employees in the same way. A symbolic award which is not really accepted among the workforce, because it does not really reflect good performance, or which is not intentioned sincerely by a supervisor, could backfire. The size of peer groups and the degree of publicity, expressed by the hierarchy levels in our experiments, are crucial for the success of the trophy value awards, which do not have a monetary value of their own. Therefore, we advise not being extravagant with these gestures and also being aware that unwanted implications of this kind of tournament might also arise if team work is an essential part of performance.

V. References

Agrawal, A., Knoeber, C. (1996): Firm Performance and Mechanisms to Control Agency Problems between Managers and Shareholders. The Journal of Financial and Quantitative Analysis. Vol. 31, No. 3, pp. 377-397.

Auriol, E., Renault, R. (2008): Status and Incentives. The RAND Journal of Economics. Vol. 39, Vol.1, pp. 305-326.

Ball, S., Eckel, C. (1998): The Economic Value of Status. Journal of Socio-Economics. Vol. 17.4, pp. 495-514.

Ball, S., Eckel, C., Grossman, P.J. and Zame, W. (2001): Status in Markets. Quarterly Journal of Economics. Vol. 155.1, pp. 161-181.

Blanes i Vidal, J., Nossol, M. (2009): Tournaments Without Prizes: Evidence from Personnel Records. Working Paper available at SSRN.

Clark, A., Masclet, D. and Villeval, M. (2010): Effort and Comparison Income: Experimental and Survey Evidence. Industrial & Labor Relations Review. Vol. 63.1, pp. 407-426.

Delfgauuw, J., Dur, R., Sol, J., Verbeke, W. (2009): Tournament Incentives in the Field: Gender Differences in the Workplace. Tinbergen Institute Discussion Paper (TI 2009-069/1).

Dur, R. (2009): Gift Exchange in the Workplace: Money or Attention? European Economic Association. Vol.7, No. 2-3, pp. 550-560.

Egon Zehnder International (2009): 6th International Executive Panel.

Ellingsen, T., Johannesson, **M**. (2007): Paying Respect. Journal of Economic Perspectives. Vol. 21,No 4, pp. 135-149.

Ellingsen, T., Johannesson, M. (2008): Pride and Prejudice: The Human Side of Incentive Theory. American Economic Review. Vol. 98.3, pp. 990-1008.

Eriksson, T., Poulsen, A., Villeval, M. (2009): Feedback and incentives: Experimental evidence. Labour Economics. Vol.16, pp. 679-688.

Falk, A., Ichino, A. (2006): Clean Evidence on Peer Pressure. Journal of Labor Economics, Vol.24, no.1, pp. 39-58.

Frey, B. (1997): Not just for the money. An Economic Theory of Personal Motivation. Edward Elgar Publishing Limited.

Frey, B., Neckermann, S. (2009): Awards in Economics Towards a New Field of Inquiry. CESifo Working Paper No. 2515.

Frey, B., Neckermann, S. (2010): Awards as Signals. CESifo Working Paper No. 3229.

Fischbacher, U. (2007): z-Tree: Zurich Toolbox for Ready-made Economic Experiments, Experimental Economics. Vol. 10.2, pp. 171-178.

Gneezy, U., Rustichini, A. (2000): Pay enough or don't pay at all. Quarterly Journal of Economics. Vol. 115, pp. 791-810.

Jeffrey, S.A. (2009): Justifiability and the motivational power of tangible noncash incentives. Human Performance. Vol. 22, pp. 143-155.

Kosfeld, M., Neckermann, S. (2010): Getting More Work for Nothing? Symbolic Awards and Worker Performance. IZA Discussion Paper No. 5040.

Lazear, E. (1998): Personnel economics for managers. John Wiley &Sons. Chapter: Nonmonetary Compensation pp. 377-409.

Lazear, E. (2000): Performance Pay and Productivity. American Economic Review. Vol. 90. pp. 1346-1361.

Mohnen, A., Pokorny, K. and Sliwka, D. (2008): Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence. Journal of Labor Economics. Vol. 26. pp. 693-720.

Moldovanu, B., Sela, A. and Shi, X. (2007): Contest of Status. Journal of Political Economy. Vol. 115.2, pp. 338-363.

Neckermann, S., Frey, B. (2008): Awards as Incentives. Working Paper Series ISSN 1424-0459 Univeristy of Zurich.

Neckermann, S., Cueni, R. and Frey, B. (2009): What is an award worth? An Econometric Assessment of the Impact of Awards on Employee Performance. CESifo Working Paper No. 2657.

Niederle, M., Vesterlund, L. (2007): Do Women Shy Away from Competition? Do Men Compete Too Much? The Quarterly Journal of Economics, MIT Press, Vol. 122(3), pp. 1067-1101.

Pokorny, K. (2008): Pay – but do not pay too much. An experimental study on the impact of incentives. Journal of Economic Behavior and Organization. Vol. 66, pp. 251-264.

Shaffer, V., Arkes, H. (2009): Preference reversals in evaluations of cash versus non-cash incentives. Journal of Economic Psychology, Vo.30 (6), pp. 859-872.

Sorauren, I. F. (2000): Non-Monetary Incentives: Do people work only for money? Business Ethics Quarterly. Vol.10.4, pp. 925-944.

VI. Appendix

Table 1: Descriptive Variables

Variable	Mean	Std. Dev.	Min	Max	Description
score1	41.51	17.85	0.00	88.00	Numbers of correct solutions in rounds one, two, and three, respectively.
score2	49.93	23.72	0.00	98.00	
score3	54.03	28.63	0.00	110.00	
percent1	90.19	18.75	0.00	100.00	Percentage of correct solutions to worked equations in each round.
percent2	86.53	23.59	0.00	100.00	
percent3	83.16	29.11	0.00	100.00	
outtime1	97.68	159.20	0.00	616.25	Time spent on article reading (in seconds).
outtime2	72.08	167.53	0.00	625.70	
outtime3	97.68	208.30	0.00	624.69	
test time	124.13	98.63	34.86	600.98	Time needed to solve the test equations.
test error	2.13	6.61	0.00	55.00	Number of wrong answers during test stage.
male	0.64	0.48	0.00	1.00	Dummy equal 1 if participant is male.
age	26.14	5.62	21.00	61.00	Age of subjects in years.
overconfidence	0.36	0.48	0	1	Dummy equal 1 if subject assumes him/herself to be in the 1st or 2nd group whereas he or she is actually ranked in the 3rd or 4th group (round 1).
being the best	0.48	0.50	0.00	1.00	Dummy equal 1 if subject wants to be the best performer.
being not the worst	0.24	0.43	0.00	1.00	Dummy equal 1 if subject does not want to be the worst performer.

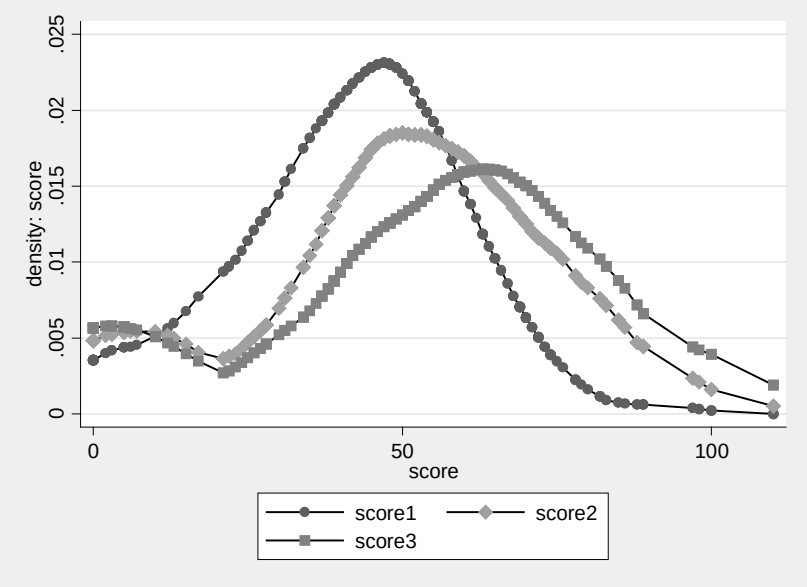

Figure A1: Densities of solved equations in round one, two, and three

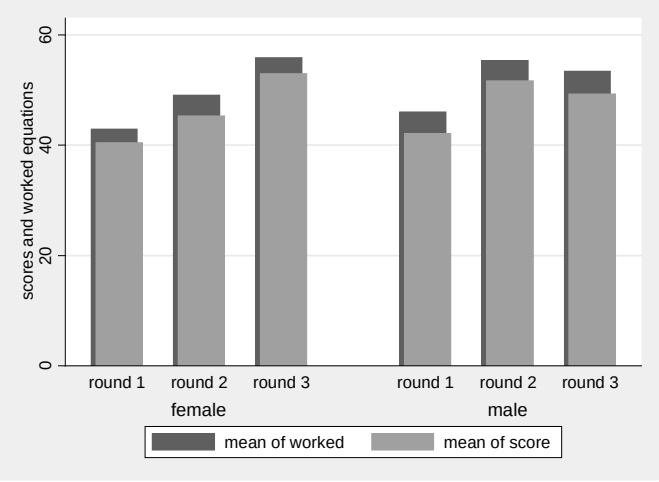

Figure A2:
Gender differences in work performance

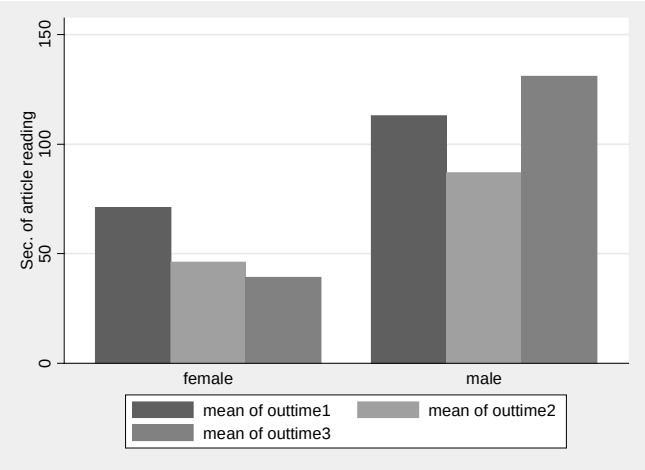

Figure A3:
Gender differences in usage of outside options

Table 3: OLS regression with differences in scores as dependent variables

Dependent variable	**Score2- Score1**		**Score3-Score2**		**Score3-Score1**	
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
test time	-0.011	0.01	-0.031**	-0.032*	-0.042*	-0.022
	(-0.47)	(-0.57)	(2.02)	(-1.94)	(-1.68)	(-1.09)
test error	-0.087	0.074	0.232	0.282	0.146	0.356**
	(-0.29)	(-0.32)	(-1.12)	(-1.29)	(-0.71)	(2.03)
award	-5.496	0.784	1.944	0.832	-3.552	1.615
	(-0.99)	-0.13	(-0.53)	(-0.2)	(-0.5)	(-0.22)
smallgroup	-10.461*	-9.222*	-2.068	-2.19	-12.529	-11.412*
	(-1.74)	(-1.82)	(-0.51)	(-0.54)	(-1.62)	(-1.68)
smallaward	14.481*	11.558	3.716	3.625	18.198*	15.183*
	(-1.73)	(-1.63)	(-0.62)	(-0.62)	(-1.74)	(-1.7)
age	-0.176	0.208	-0.126	-0.082	-0.302	0.125
	(-0.76)	(-0.7)	(-0.64)	(-0.35)	(-1.02)	(-0.33)
male	6.025	7.718*	-7.44**	-7.172**	-1.415	0.546
	(-1.39)	(-1.89)	(2.39)	(2.16)	(-0.31)	(-0.12)
overconfidence	6.501	5.763	-7.461**	-7.151**	-0.96	-1.388
	(-1.54)	(-1.52)	(2.13)	(1.99)	(-0.17)	(-0.27)
awbest		-4.917		3.26		-1.657
		(-0.62)		(-0.63)		(-0.17)
being the best		21.494***		1.572		23.066***
		(3.60)		(-0.33)		(2.88)
being not the worst		8.308*		3.365		11.673**
		(-1.95)		(-1.14)		(2.30)
Constant	12.442	-15.6	17.244**	14.268	29.687***	-1.332
	(-1.43)	(-1.38)	(2.49)	(-1.55)	(3.23)	(-0.1)
Observations	240	240	240	240	240	240
R-squared	0.1	0.32	0.17	0.19	0.08	0.28

Robust t-statistics in parentheses; significant at 10% (*), 5% (**) and 1% level (***). Observations are clustered for subjects. Dependent variables are the differences in scores between the three rounds.

Table 5: OLS regression differentiated by ranking groups

	Score2		Score3	
Dependent variable	**35% Best**	**65%Worst**	**35% Best**	**65%Worst**
score1	1.425***	0.469		
	(4.51)	(-1.64)		
score2			0.939***	0.901***
			(5.22)	(6.34)
test time	-0.218***	-0.041	-0.036	-0.043*
	(3.09)	(-1.33)	(-0.28)	(-1.7)
test error	-2.321**	-0.041	-1.427	0.348
	(2.39)	(-0.16)	(-0.63)	(-1.41)
award	2.66	-5.425	-5.816	0.563
	(-0.28)	(-0.7)	(-0.71)	(-0.08)
smallgroup	-31.31***	0.072	-8.424	-2.61
	(2.87)	(-0.01)	(-1.25)	(-0.35)
smallaward	36.697**	7.425	7.728	4.336
	(2.71)	(-0.87)	(-1.13)	(-0.38)
awbest	-7.628	-0.101	2.514	6.867
	(-0.59)	(-0.01)	(-0.32)	(-0.71)
age	-0.401	0.274	0.514	-0.191
	(-0.49)	(-0.94)	(-0.9)	(-0.62)
male	3.01	8.332*	2.563	-9.606
	(-0.77)	(-1.81)	(-0.52)	(-1.61)
overconfidence		-2.406	-9.424**	-5.575
		(-0.38)	(2.73)	(-1.03)
being the best	19.531*	25.041***	-1.846	1.356
	(-2.05)	(3.90)	(-0.26)	(-0.12)
being not the worst	6.485	10.913*	-0.545	8.016
	(-0.99)	(-1.8)	(-0.12)	(-1.47)
Constant	0.882	11.541	8.621	21.118
	(-0.03)	(-0.78)	(-0.39)	(-1.34)
Observations	30	50	30	50
R-squared	0.78	0.63	0.84	0.66

Robust t-statistics in parentheses; significant at 10% (*), 5% (**) and 1% level (***), The performance groups are a cluster of ranking groups one and two as well as three and four
Dependent variables are the absolute score values.

Instructions of the Award Treatment (translated into English)

First round:

Instructions:

Dear Participant,

The experiment consists of three rounds. Each round lasts **ten minutes.** You can either solve simple mathematical equations or read articles from different genres (See button next to equations). After filling in your answers, you can switch to the next task screen by pressing "Continue" at the bottom of each page. If you choose to read articles, you can at any time return to the equation screens by clicking on the button "Back".

The program will stop automatically after 10 minutes. You will then get the instructions for the next round.

The results of each round are not connected with each other. Furthermore, your payoff of **15 Euros** at the end is **independent** of your behavior during the experiment**.**

Test Task:

Each equation contains of two rows with three one-digit numbers. You have to add up each row and afterwards subtract the lower number from the higher one. You only have to fill in the final solution.

Example:

7+3-8=2
8+9-8=9
Solution= 7

On the next slide you will find five simple test equations. Here you can check whether you have understood the task instructions correctly.

You can only carry on if all five test equations have been solved correctly.

Please confirm your having understood by clicking on the button „To the Experiment". If any questions arise, please give a hand signal to an instructor. The program will continue automatically and simultaneously when all of you have clicked on this button.

Task Screen Round One: (see Figure 1)

Self-Assessment:

Assess your performance of the first round by marking one of the five categories below. If your self-assessment is right, you will get an additional payoff of 2 Euros at the end of the experiment. Whether your guess was right or not will be disclosed at the end. There will be no further bonus questions during the experiment.

I rank among the... (With regard to the number of correct answers compared to those of the other participants)

❑ Best 5% (Ranking position: 1)
❑ 30% (Ranking position: 2-8)
❑ 45% (Ranking position: 9-18)
❑ Worst 20% (Ranking position19-22)

Second Round:

Instructions:

Your task in the **second round** is similar to the first one. Again you have the choice between solving simple mathematical equations and reading articles from different genres. You have to click on the button "Articles" to read (see button next to equations) and by pressing "Back" you can switch back to the equation screen.)

At the end of this round you will get an assessment of your performance with information about the number of solved equations and your ranking position compared to those of the other participants. The ranking position is measured by the number of correct answers and in the case of two participants solving an equal number of equations correctly, by the proportion of correct answers over all answers given.

The results are given as followed:

- Number of given answers:
- Percentage of correct answers:
- Ranking: You rank among the:

    - Best 5% (Ranking position: 1)
    - 30% (Ranking position: 2-8)
    - 45% (Ranking position: 9-18)
    - Worst 20% (Ranking position19-22)

At the end of the experiment you will get an award for your performance in the form of gold, silver, or bronze buttons according to your ranking position. The award is equally likely to be based on the performance of either the second or the third round. The following illustration shows the awards.

Illustration of the Awards:

Task Screen Round Two: (see Figure 1)

Result Screen:

- Number of given answers:
- Percentage of correct answers:
- Ranking: You rank among the:

    - Best 5% (Ranking position: 1)
    - 30% (Ranking position:2-8)
    - 45% (Ranking position: 9-18)
    - Worst 20% (Ranking position19-22)

Third Round:

Instructions:

The task in the **third round** is similar to the tasks before. Again, you have the choice of solving simple mathematical equations or reading articles from different genres.

At the end of this round, you will get an assessment of your results including the number of worked and solved equations and your ranking position compared to those of all other participants. The **ranking position** again depends on the **number of correct answers** or, if two participants have the same number of correctly solved equations, the proportion of correct answers over all answers given.

The decision about which round is relevant for the award is random. Important is always your ranking position.

Task Screen Round Three: (See Figure 1)

Result Screen:

- Number of given answers:
- Percentage of correct answers:
- Ranking: You rank among the:
    - Best 5% (Ranking position: 1)
    - 30% (Ranking position: 2-8)
    - 45% (Ranking position: 9-18)
    - Worst 20% (Ranking position19-22)

Payoff Screen:

Your payoff: 15.00 Euros

- Worked/solved equations (round 1) : 0/0
- Was self-appraisal round 1 correct? : No/Yes
- Worked/solved equations (round 2) : 0/0
- Worked/solved equations (round 3) : 0/0

Please remain seated after finishing the questionnaire until your place number is called out. You will then receive your payoff of 15 Euros or - if your self-assessment was correct - 17 Euros.