# Online Supplementary Material

Accompanying the research article

"Work Motivation and Teams"

by

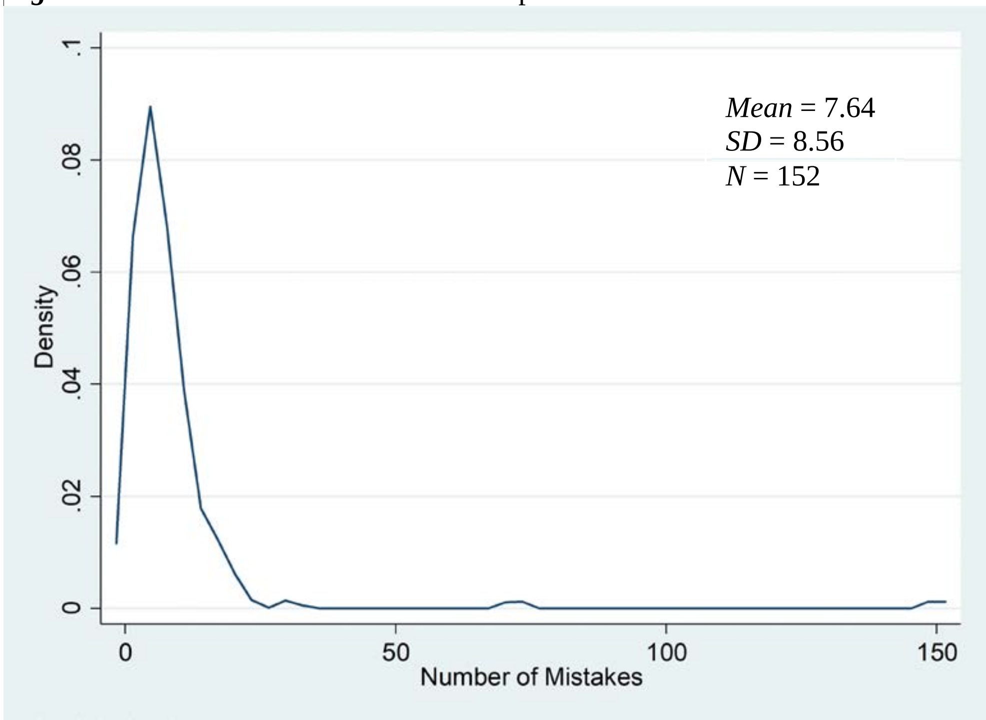
Simone Haeckl, Rupert Sausgruber, and Jean-Robert Tyran

August 15th 2018

# I. Robustness checks

## a. Control for mistakes

***Figure OSM 1:*** Distribution of mistakes in phase 1




*Notes*: Kernel density function of the number of mistakes over all treatments.

As mentioned in section 1 of the main paper, we counted the number of mistakes, defined as how often a worker submits incorrect cross sums, made by each worker as a robustness check. A high number of mistakes could either be caused by participants having trouble with the math task or not taking the experiment seriously. Therefore, we now test the robustness of our results to only considering participants who are likely to understand the math task and work conscientiously. First, we will give a short overview of the distribution of mistakes. Second, we show that excluding workers who made more than 10 mistakes (which corresponds to about 10% of the average output) does not change the results of our main regressions qualitatively.

*Figure OSM 1* shows the kernel density function for the number of mistakes. On average workers make 8 mistakes (sdev = 12.69), only 2 workers make more than 20 mistakes and 3 workers make no mistakes. In tables *OSM 1* and *OSM 2*, we replicate the main results in the paper after excluding 45 workers who made more than 10 mistakes. *Table OSM 1* repeats the regression of *Table B1* for output as the dependent variable. It replicates the result that in the Team treatment output increases compared to the Individual treatment (see the coefficient *Phase 2 x Team* in column (1): 19.84 and compare it to the equivalent coefficient in *Table B1*: 18.43). In line with the findings in *Table B1* this effect is mainly driven by team incentives and not by observation as we find no significant effect for *Phase 2 x Info* in column (1) but a significant increase for variable *Phase 2 x Team* in column (2) (in *Table B1* the coefficients are 4.53, $p$ = 0.215, for *Phase 2 x Info* and 13.90, $p$ = 0.001, for *Phase 2 x Team*).

***Table OSM 1:*** Workers in Team increase output significantly in phase 2

| | Output | |
|---|---|---|
| | (1) | (2) |
| | Baseline: Individual | Baseline: Info |
| Phase 2 | 3.79 | 8.64* |
| | (2.77) | (4.48) |
| Info | -2.57 | |
| | (7.53) | |
| Team | -6.23 | -3.66 |
| | (6.49) | (6.68) |
| Phase 2 x Info | 4.84 | |
| | (5.25) | |
| Phase 2 x Team | 19.84*** | 15.00*** |
| | (3.99) | (5.32) |
| _cons | 89.21*** | 86.64*** |
| | (5.22) | (5.45) |
| $R^2$ | 0.09 | 0.11 |
| Subjects | 108 | 74 |
| Observations | 216 | 148 |

*Notes*: OLS regression with standard errors clustered by team. Column (1) includes all treatments, in column (2) only the Info and Team treatment are included. We had a total of 51 subjects participating in each treatment and, as each subject appears in phase 1 and phase 2, this would lead to 306 observations in column (1) and 204 observations in columns (2). However, 45 (28 for column 2) subjects are omitted as they made at least 10 mistakes or completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Numbers in parentheses indicate clustered robust standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

*Table OSM 2* repeats the regression of *Table B4* for output as the dependent variable. We investigate the moderating effects of motivation on team incentives and observation, and again find that motivation moderates team effects controlling for information (see variable *Team x Negative Motivation* in column 2).

***Table OSM 2:*** Motivation moderates team incentives and observation

| | Change in Output | |
|---|---|---|
| | (1)<br>Baseline: Individual | (2)<br>Baseline: Info |
| Info | -1.60 | |
| | (13.08) | |
| Team | 1.40 | 3.00 |
| | (7.20) | (11.92) |
| No Motivation | 11.38 | 24.60* |
| | (7.77) | (12.33) |
| Negative Motivation | 1.25 | 0.85 |
| | (6.96) | (13.36) |
| Info x No Motivation | 13.22 | |
| | (14.56) | |
| Info x Negative Motivation | -0.40 | |
| | (15.05) | |
| Team x No Motivation | 22.41** | 9.19 |
| | (10.27) | (14.05) |
| Team x Negative Motivation | 28.95*** | 29.35* |
| | (8.85) | (14.43) |
| _cons | -2.00 | -3.60 |
| | (6.37) | (11.44) |
| $R^2$ | 0.39 | 0.40 |
| *N* | 108 | 74 |

*Notes*: OLS regression with standard errors clustered by team. Column (1) includes all treatments, in column (2) only the Info and Team treatment are included. Positively motivated subjects are used as the baseline. However, 45 (28 for column 2) subjects are omitted as they made at least 10 mistakes or completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Numbers in parentheses indicate standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

**Truncated Dataset.** A possible driver of our results is a difference in the accuracy of the estimated cost function eq. (3) based on output in phase 1. For workers who stop working early on the real-effort task, there are fewer observations available to estimate their cost functions, which could bias the estimation results. In particular, as the task gets more difficult over time, the estimated parameters may not capture the actual curvature of the cost function for low productive or negatively motivated workers. This may bias upward the estimation of the money-maximizing output and consequently our measure of motivation (i.e., negatively motivated workers look more negatively motivated). To investigate whether our results are robust to such potential bias, we trim the available data for high-productive subjects and repeat the estimation of the cost functions with data truncated from above. In "truncated definition 1" we exclude the last 9 tasks (10% of the average output) from the estimation of the cost function for high-productive participants (defined by median split). In "truncated definition 2" we exclude the last 18 tasks (20% of the average output) for those subjects, respectively.

***Table OSM 3:*** Motivation moderates team incentives and observation (using truncated data)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Full dataset | Tuncated: Definition 1 | Tuncated: Definition 2 |
| Info | 3.19 | 1.88 | 1.50 |
| | (5.26) | (5.45) | (4.70) |
| Team | 4.13 | 2.81 | 8.63 |
| | (4.29) | (4.52) | (5.72) |
| Negative Motivation | 12.50** | 11.18* | 13.79** |
| | (5.73) | (5.92) | (5.90) |
| Money Maximizer | 2.79 | 0.24 | 4.26 |
| | (4.85) | (4.92) | (4.50) |
| Info x Negative Motivation | 7.58 | 6.07 | 7.85 |
| | (8.43) | (9.78) | (8.45) |
| Info x Money Maximizer | -4.10 | 1.44 | 0.81 |
| | (7.57) | (6.82) | (6.04) |
| Team x Negative Motivation | 19.00** | 20.32** | 12.84 |
| | (8.04) | (8.17) | (9.35) |
| Team x Money Maximizer | 24.21*** | 26.76*** | 16.93** |
| | (6.05) | (6.11) | (8.03) |
| _cons | -2.44 | -1.13 | -3.38 |
| | (3.78) | (4.04) | (3.67) |
| $R^2$ | 0.39 | 0.36 | 0.34 |
| *N* | 147 | 147 | 147 |

*Notes*: OLS Regression with robust standard errors clustered by teams using a diff-in-diff approach to identify interaction effects between individual motivation and observation as well as team incentives. Negatively motivated subjects are used as the baseline. Six subjects are omitted as they made at least 10 mistakes or completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Numbers in parentheses indicate standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

*Table OSM 3* replicates our main result using the truncated data. Naturally, the estimate for the money-maximizing output gets less precise (see decrease in $R^2$ from column 1-3) and the explanatory power of our motivation measure decreases the fewer observations we use for estimating the cost functions. However, the results stay qualitatively the same as for the full data set.

## b) Replication of results with reference to optimal stopping time

**Average deviation per screen.** As our measure of motivation $\Delta Q$ depends on $Q^*$ which is based on the cost function estimated for each participant (see eq. 3), we want to introduce another measure for motivation, independent of our estimation as a robustness check. Therefore, we measure how much the workers deviate from 31.5 seconds, the point at which wage per screen equals the monetary opportunity-cost of work, on each screen. To do so, we sum up the net deviations from 31.5 seconds, starting at the screen before the worker for the first time has taken more than 31.5 seconds to complete a screen. For workers who never took more than 31.5 seconds to complete a screen, we use the time they have spent to complete the last screen. For simplicity, we call this measure aggregated deviation per screen (*ADS*). *Figure OSM 2* illustrates the concept. *ADS* is equivalent to the highlighted area.

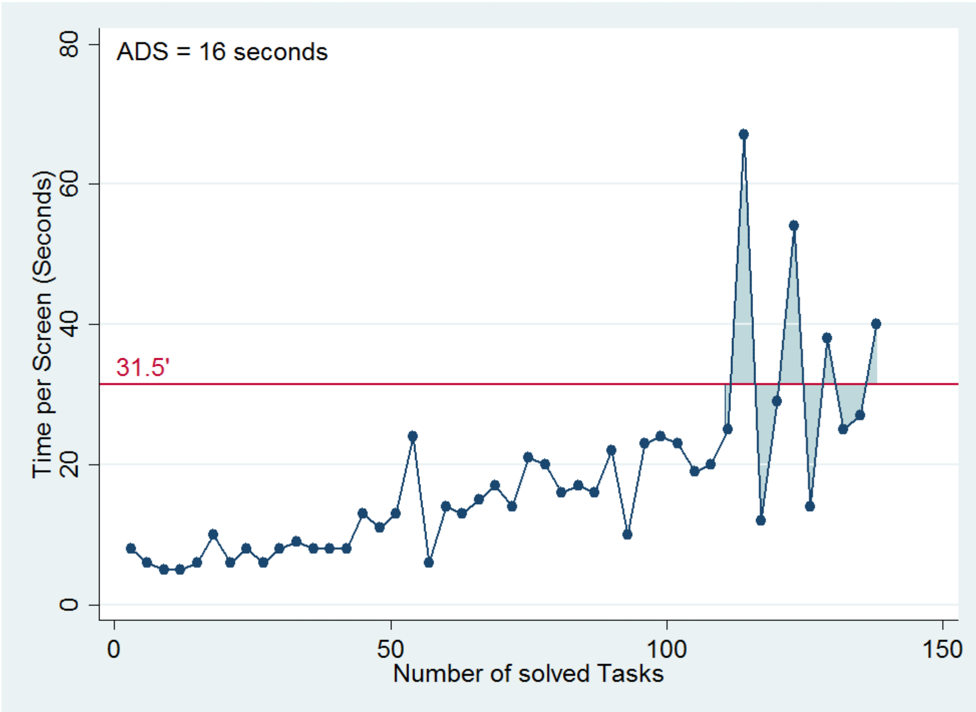
The areas above the 31.5 seconds line are additional seconds spent on a screen, highlighted areas below the 31.5 seconds line are subtracted from the additional time in order to receive a measure of net deviation from 31.5 seconds. This is outlined in equation (1), in which $\tau_j$ represents the time it takes worker $i$ to solve the $j^{th}$ screen. As shown in *Figure OSM 2*, the aggregation begins one screen before the worker for the first time has taken more than 31.5 seconds to solve one screen $((\min_j \tau_j > 31.5) - 1)$.

$$ADS_i = \sum_{j=\min(j)|(\tau_j>31.5)-1}^{J} \max_\tau(\tau_{ij} - 31.5,0) - \max_\tau(31.5 - \tau_{ij}, 0) \tag{1}$$

Now, we first describe *ADS* and second replicate the main result we obtained with $\Delta Q$ for *ADS*, in order to show that our results are robust. This is an important step, as it supports the claim that our results are not caused by a biased measure of money-maximizing output.

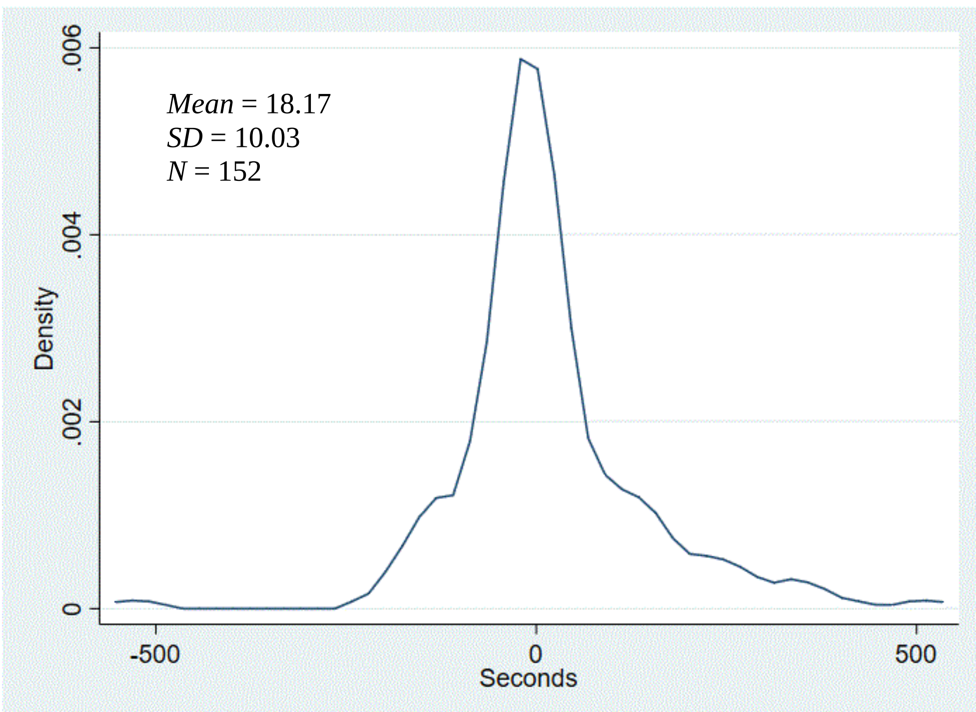
*Figure OSM 3* shows the kernel density for *ADS* in phase 1 using the data of the Individual treatment. Using this measure for money-maximizing output, workers are on average positively motivated in phase 1 and have an *ADS* of 18.17 seconds ($p$ = 0.658, $t$-test).

***Figure OSM 2:*** Time per screen




*Notes*: Visualization of the deviation variable in a time per screen over number of completed tasks space.

***Figure OSM 3:*** Aggregate deviation per screen




*Notes*: Kernel density estimate of aggregate deviation per screen in phase 1. One subject is omitted as he or she did not solve any tasks in phase 1.

*Table OSM 4* shows an OLS regression with dummies for motivation types based on *ADS* in phase 1. We find qualitatively the same results as for our motivation measure $\Delta Q$ (see *Table B4*).

***Table OSM 4:*** Motivation (measured as *ADS*) moderates team incentives and observation

| | Change in Output | |
|---|---|---|
| | (1) | (2) |
| | Baseline: Individual | Baseline: Info |
| Info | -0.71 | |
| | (4.54) | |
| Team | 6.47 | 7.17 |
| | (5.84) | (5.67) |
| Negative Motivation | 3.53 | 9.53 |
| | (4.95) | (5.67) |
| Money Maximizer | 4.32 | 14.73* |
| | (4.88) | (7.44) |
| Info x Negative Motivation | 6.00 | |
| | (7.51) | |
| Info x Money Maximizer | 10.41 | |
| | (8.87) | |
| Team x Negative Motivation | 17.95* | 11.95 |
| | (9.26) | (9.68) |
| Team x Money Maximizer | 13.39* | 2.98 |
| | (7.74) | (9.58) |
| _cons | 0.18 | -0.53 |
| | (3.38) | (3.04) |
| $R^2$ | 0.23 | 0.22 |
| $N$ | 146 | 96 |

*Notes*: OLS Regression using a diff-in-diff approach to moderating effects of motivation on the effectiveness of team incentives and observation by peers. Positively motivated subjects are used as the baseline. As in the main analysis six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Additionally, one subject is omitted who apparently has stopped working without switching to leisure (among all subjects, this happened just once and presumably by mistake). Numbers in parentheses indicate clustered robust standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

## b. Robustness regarding alternative definitions of motivation groups

While using terciles to split the observations into different motivation groups is very convenient as it generates equal group sizes, one could also think of different ways to define negative or positive motivation types. *Table OSM 5* replicates *Table B4* using two additional definitions of motivation groups.

***Table OSM 5:*** Moderating results are qulitatively robust to changes in the definition of motivation

| | Change in Output | |
|---|---|---|
| | (1)<br>Motivation Definition 1 | (2)<br>Motivation Definition 2 |
| Info | 1.22 | 0.29 |
| | (4.57) | (4.48) |
| Team | 5.18 | 9.83* |
| | (4.13) | (5.19) |
| Negative Motivation | 15.68*** | 18.45*** |
| | (4.69) | (6.23) |
| Money Maximizer | 3.14 | 4.13 |
| | (4.44) | (3.79) |
| Info x Negative Motivation | 11.72 | 11.84 |
| | (8.89) | (10.25) |
| Info x Money Maximizer | 3.59 | 6.48 |
| | (5.79) | (6.85) |
| Team x Negative Motivation | 15.15** | 9.74 |
| | (7.43) | (9.78) |
| Team x Money Maximizer | 19.67*** | 13.80** |
| | (5.96) | (6.75) |
| _cons | -2.82 | -2.25 |
| | (3.51) | (3.03) |
| $R^2$ | 0.37 | 0.38 |
| *N* | 147 | 147 |

*Notes*: OLS Regression with cluster robust standard errors using a diff-in-diff approach to identify interaction effects between individual motivation and observation by peers as well as team incentives. Positively motivated subjects are used as the baseline. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Numbers in parentheses indicate clustered robust standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

In column (1), we define Money Maximizers as participants deviating by less than half a standard deviation of $\Delta Q$ from the mean $\Delta Q$. As this definition is based on the mean deviation form money-maximizing output, it identifies motivation relative the other workers. We find qualitatively the same results as in Table B4 (see *Team x Negative Motivation* and *Team x Money Maximizer*). In column (2), we define Money Maximizers as participants deviating by less than half a standard deviation of $\Delta Q$ from 0. We replicate our result for Money Maximizers, but while the coefficient for negatively motivated workers has still the expected sign, it is no longer significant. This can be explained by the unbalanced group size which results from using this definition (36 participants are negatively motivated, while 66 are positively motivated).

## II. Are there gender differences?

Especially in the Info treatment, we expected to find Gender effects as there is an extensive literature discussing differences in competitive behavior. Whiles most of the literature finds that men increase performance more strongly in tournament settings or settings in which the sex of the competitors was made salient, results are less clear in other incentives schemes (see for example Bellemare et al. 2010, Gneezy, Niederle and Rustichini 2003, Niederle and Vesterlund 2007, Niederle and Vesterlund 2008).

*Table OSM 6* replicates the OLS regression analysis used to identify treatment effects on output (see *Table B1*, column 2) for males and females separately (see columns 1 and 2), and reports differences between the gender specific coefficients (see column 3, *p*–values are based on Chow tests). We find a significant effect of the Team treatment for both genders (see *Phase2 x Team* in columns 1 and 2). This effect does not differ between men and women ($p$ = 0.189).

As expected, we find that the Info treatment effects men and women differently. In phase 1, males solve 13.50 tasks more in Info than in Individual ($p$ = 0.084), while there is no difference for females (-5.70 tasks, $p$ = 0.410). This adds up to a gender-effect of 19.20 tasks for Info in Phase 1 ($p$ = 0.058, see variable *Info*). This indicates that knowing that the performance will be announced increases the effort provided by men and is in line with a stronger reaction to the announcement of feedback by males (Kuhnen and Tymula 2012).

In phase 2 (note that we provided the participants with feedback on relative performance between phases in Info), females increase output by 12.70 tasks more in Info than in Individual ($p$ = 0.023), while males do not increase output due to the performance feedback (-2.51 tasks, $p$ = 0.471), adding up to a gender difference of -15.21 tasks ($p$ = 0.082). This gender effect could be caused by a difference in satisfaction with one's performance between males and females. Literature has shown that women tend to be less satisfied with their performance, even if they achieve the same results as men (Deaux and Farris 1977). Not being satisfied with one's own performance would be a good incentive to deviate from money-maximizing output in phase 2. Using a self-reported satisfaction measure, we see that non-satisfied workers increase the distance to money-maximizing output significantly more than satisfied workers (21 vs. 1 tasks, $p = 0.035$, Wald test). Looking at the distribution of this satisfaction measure over gender shows that two thirds of the non-satisfied workers are female while satisfied workers are equally divided over gender ($p = 0.157$, two-sided Fisher's exact test).

***Table OSM 6:*** Regression including Treatments and Gender Effects

| | Output | | |
|---|---|---|---|
| | (1)<br>Male | (2)<br>Female | (3)<br>Difference (1)-(2) |
| Phase 2 | 5.81** | -2.67 | 8.48** |
| | (2.59) | (3.06) | (3.75) |
| Info | 13.50* | -5.70 | 19.20* |
| | (7.65) | (6.85) | (9.99) |
| Team | -3.64 | -12.92* | 9.27 |
| | (7.62) | (7.50) | (10.96) |
| Phase 2 x Info | -2.51 | 12.70** | -15.21** |
| | (3.46) | (5.38) | (6.18) |
| Phase 2 x Team | 15.29*** | 23.97*** | -8.67 |
| | (4.29) | (4.96) | (6.51) |
| _cons | 95.25*** | 91.67*** | 3.58 |
| | (5.42) | (4.24) | (6.00) |
| $R^2$ | 0.09 | 0.07 | |
| Subjects | 80 | 67 | 147 |
| Observations | 160 | 134 | 294 |

*Notes*: OLS regression with standard errors clustered by group for output. The sample includes 153 (84 males and 69 females) subjects. As each subject appears twice (in phase 1 and phase 2), this would lead to 306 (168 males and 138 females) observations. However, six subjects (4 males and 2 females) are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Numbers in parentheses indicate clustered robust standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

## References


Bellemare, C., Lepage, P., Shearer, B. 2010. Peer Pressure, Incentives, and Gender: An Experimental Analysis of Motivation in the Workplace. *Labour Economics*. **17**(1) 276-283.

Deaux, K., & Farris, E. 1977. Attributing causes for one's own performance: The effects of sex, norms, and outcome. *Journal of Research in Personality*. **11**(1) 59-72.

Gneezy, U., Niederle, M., Rustichini, A. 2003. Performance in Competitive Environments: Gender Differences. The Quarterly Journal of Economics. **118**(3) 1049-1074.

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

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

Niederle, M. and Vesterlund, L. 2008. Gender Differences in Competition. *Negotiation Journal*. **24** 447-463.