Discussion Papers
Department of Economics
University of Copenhagen

No. 18-08

Work Motivation and Teams

Simone Haeckl, Rupert Sausgruber and Jean-Robert Tyran

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Work Motivation and Teams

Simone Haeckl, Rupert Sausgruber and Jean-Robert Tyran*

August 16th, 2018

Abstract

We provide a new measure of work motivation and show that motivation shapes the effects of team incentives and observation by peers on performance. In particular, we measure motivation to work hard as the deviation from the money-maximizing benchmark in a real-effort experiment. While we find that average output increases in response to team incentives and observation, we find that highly motivated workers do not respond. The reason is that highly motivated workers already work hard and increasing effort even further is very costly to them.

JEL code: C91, J33, L20

Keywords: real-effort experiment, cooperation, team, intrinsic motivation, labor

---

* Haeckl: Department of Economics, Vienna University of Economics and Business, Institute for Public Sector Economics, Vienna, Austria. Sausgruber: Department of Economics, Vienna University of Economics and Business, Institute for Public Sector Economics, Vienna, Austria. Tyran: Department of Economics, University of Vienna, 1090 Vienna, Austria and Department of Economics, University of Copenhagen, Copenhagen, Denmark and CEPR London. All three authors are also member of the Vienna Center of Experimental Economics at the University of Vienna. We gratefully acknowledge financial support by the Austrian Science Fund (FWF), under Project no. I 2027-G16.

1 Introduction

Why do some people work hard while others do not? This is an important question for economics and management and the economist's answer is disarmingly simple: because people optimize and face different costs and benefits of work. However, this simple principle is hard to demonstrate in practice, because the costs and benefits are often difficult to measure. For example, assume two people are offered to work on the same task for the same wage, and assume that one works harder than the other. In general, we cannot know whether this observation is in line with the optimization principle of standard economics. The reason is that the two workers might have different effort cost or opportunity cost of work and these costs are hard to observe empirically. In addition, work may have a number of non-material benefits as the literature in work psychology has pointed out. Work may give us an identity, a positive self-image, satisfaction to be good at something, pleasure to serve others, or a positive reputation in the community (Kanfer, Chen and Pritchard 2008). Importantly, work is often intrinsically enjoyable or boring in varying degrees to workers (see Della Vigna and Pope 2016).

This paper proposes a novel experimental design to measure intrinsic motivation to work hard, defined as how intrinsically enjoyable or boring it is to work relative to consuming leisure, net of effort costs. To do so, we measure deviations from optimal labor supply *for a money-maximizing* worker, i.e., a worker who is not subject to psychological cost or benefits of work and leisure beyond the purely monetary cost and returns. We control monetary returns of both work and leisure (i.e., the monetary equivalent of the opportunity cost of work) in a clean setting in which we can exclude economic and social interaction and thus reputation effects as well as the effects of social preferences. In particular, we let workers choose between work on a real-effort task and leisure. We control for the monetary opportunity cost of work by paying workers per second they spend in leisure. Workers are paid a piece rate, but the real-effort task becomes more demanding over time. Hence, the monetary opportunity cost per task increases over time. A money-maximizing worker will work when the monetary opportunity cost of work is low and switch to leisure when it is sufficiently high. Therefore, our design allows us to determine the optimal labor supply (or, equivalently, the optimal output) for a money-maximizing worker, i.e., for a worker without any non-monetary costs or benefits of work or leisure.

The optimal output of a money maximizer provides the benchmark to measure his or her intrinsic motivation to work hard. Specifically, we measure intrinsic motivation as the deliberate deviation from optimal output, controlling for other reasons to deviate like error and confusion. To prevent confusion as far as possible, we saliently inform participants in our experiment when to optimally switch to leisure, i.e., how to maximize their earnings. The aim of this unusual experimental procedure is to make sure that deviations from money maximization are deliberate rather than accidental. Furthermore, we reduce the scope for biases like overconfidence or optimism as a source of deviations from money maximization by providing participants continuously with feedback on the time it took them to complete the previous task. Our procedure seems to have been successful since observed deviations from the money-maximizing benchmark do not correlate with standard measures of cognitive ability. Previous studies (e.g., Hoppe and Kusterer 2011) have shown that cognitive ability correlates with biased choices in analytical problems with an objectively correct solution. Since our work task also has an objectively correct solution for a money maximizer (viz. when to switch to leisure to maximize earnings), and we do not find a correlation between cognitive ability and deviations from our benchmark, we conclude that switching choices (or, optimal output) were not driven by such biases.

Our design affords tight control since we provide uniform and known conditions across workers both with respect to work and leisure. This control is an important advantage over field studies that often cannot observe leisure activities of workers (e.g., Bandiera, Barankay and Rasul 2013, Boning, Ichniowski and Shaw 2007, Erev, Bornstein and Galili 1993, Hamilton, Nickerson and Owan 2003).

We use our novel experimental measure of intrinsic motivation to classify workers into those who have positive and those who have negative intrinsic motivation. We then investigate which type of worker works harder when assigned to a team rather than working alone. We study the effect of work motivation in teams because teams are of paramount importance in modern firms, and increasingly so. For example, Deloitte, one of the world's biggest accounting firms, headlined its annual report in 2016: "Organizational design - The rise of teams".

While there is a rich literature on various aspects of teamwork such as joint production (Hamilton et al. 2003), cooperation (Berger, Herbertz and Sliwka 2011), communication (Boning et al. 2007), endogenous team formation (Bandiera, Barankay and Rasul 2010), and team tournaments (Bandiera et al. 2013, Reuben and Tyran 2010), we are, to the best of our knowledge, the first to show *how heterogeneity in intrinsic motivation shapes the effectiveness of team incentives compared to individual incentives*. To test for the effect of heterogeneity in intrinsic motivation, we use a tightly controlled setting in which we focus on the two most prominent characteristics of teams: an externality between team members' payoffs, which we call (following Babcock et al. 2015) *team incentives* and the ability of team members to observe each other's output, which we call *observation*.[1] More specifically, each task completed in the Team treatment generates earnings for the other team members, and workers can observe their team members' performance while neither of those characteristics is present in the Individual treatment (a control treatment called Info, serves to isolate the effect observation).

[1] Observation can lead to peer pressure. However, evidence on the effects of peer pressure is mixed. While some authors find it improves performance (see Falk and Ichino 2006, Mas and Moretti 2009, Mohnen, Pokorny and Sliwka 2008, Tran and Zeckhauser 2012) others find no or even negative effects (see Barankay 2012, Bellemare, Lepage and Shearer 2010, Eriksson, Poulsen and Villeval 2009). See also Herbst and Mas (2015) for a review.

We are able to investigate how different types of workers respond to team incentives and observation because we implement a design in which a money maximizer behaves in the *same* way whether he or she is in a team or not. That is, the individually optimal output for a selfish money maximizer is *not* affected by externalities and observation in our design. In contrast, workers with social preferences are motivated to work harder under these conditions.

We find that team incentives and observation by peers increase output by 25 percent for the average worker. We show that this increase is mainly driven by workers with negative or no intrinsic motivation. We suggest that this is so because of cost considerations. These cost considerations come into play because our design picks up a fundamental aspect of well-behaved production processes, namely that "hills are flat at the top" of continuously differentiable concave function. In particular, we argue that expanding output is profitable for those with output below the money-maximizing output (i.e., those with negative intrinsic motivation) but expensive for those beyond the optimum (i.e., those with positive intrinsic motivation). Under the assumption that cooperativeness is equally distributed among those with positive and negative intrinsic motivation, these cost considerations imply that those with negative intrinsic motivation react more strongly to team incentives than those with positive intrinsic motivation (a follow-up experiment shows that the assumption holds, see appendix B 4).

Our paper relates to the following strands of literature. First, our paper is closely related to research on intrinsic motivation. A common definition of intrinsic motivation is that it reflects the pleasure people derive from merely exercising a task (Deci 1975). Previous attempts to measure intrinsic motivation in economics have focused on motivation crowding. Such crowding occurs when the presence of monetary incentives reduces intrinsic motivation and, as a consequence, effort provision (Bowles and Polania-Reyes 2012, Deci 1971, Frey and Jegen 2001, Gneezy, Meier and Rey-Biel 2011). Intrinsic motivation has mainly been analyzed by removing or introducing performance-related incentives (e.g., Bognanno and Huffman 2017, Hossain and Li 2014 Irlenbusch and Sliwka 2005, Segal 2012). However, when measuring intrinsic motivation rather than motivation crowding, a problem with this approach is that behavior is undetermined in the absence of incentives. In this case, economics makes no clear prediction, and confounds arising from situational aspects as well as experimenter demand effects loom large (Zizzo 2010). In contrast, we measure intrinsic motivation *in the presence of performance-related monetary incentives* which reduces the possible confounds in the measurement of intrinsic motivation, because it provides a point estimate of money-maximizing output.

Second, we contribute to the rich experimental literature on team incentives with observation. For example, Azmat and Iriberri (2016) compare the effect of observation on output under an individual piece rate and a flat payment, Sausgruber (2009) analyzes the effect of between-team observation on team production, and Georganas, Tonin and Vlassopoulos (2015) study the effect of being an observer vs. being observed. However, none of these studies features a leisure task at all. Yet, the inclusion of a leisure task can influence the effect of incentives, as discussed by Araujo et al. (2016), Corgnet, Hernan-Gonzalez and Schniter (2015a, 2015b), Dickinson (1999), Erkal, Gangadharan and Boon (2018), or Sausgruber, Sonntag and Tyran (2018). In contrast, the present paper has paid leisure, which allow us to control the monetary opportunity cost of work. A close match to our paper is Gjedrem and Kvaløy (2018) who have a leisure task, albeit an *unpaid* one. This means that they do not have a well-defined optimum for a money-maximizer and are not able to analyze deviations from money-maximizing behavior. However, they use a very systematical approach to analyze how observation affects behavior in individual vs. team-payment schemes.

Third, we contribute to the literature on public goods by studying team incentives in the guise of a positive externality between the team members' payoffs, which is typical for public goods (Nalbantian and Schotter 1997). While there are literally hundreds of experimental papers analyzing contributions to public goods in the absence of real effort (i.e., participants receive an endowment like manna from heaven and the act of contribution merely consists of choosing a number of tokens to keep vs. to invest in the public good, see Chaudhuri 2011 for survey), there are only few studies involving real efforts. Most of the exceptions involve a real effort to earn the endowment (e.g., Balafoutas et al. 2013, Cherry, Kroll and Shogren 2005, Muelbacher and Kirchler 2009). Only a handful of studies make the actual contribution effortful (Cooper and Saral 2013, Dutcher, Salmon and Saral 2015, Filiz-Ozbay and Ozbay 2014, Van Dijk, Sonnemans and Van Winden 2001). Yet, having real effort in contributions is important because people might not only differ in their ability to complete a task, but also in how much they like working on the task, which could moderate their willingness to contribute. We add to the literature on public goods because our design enables us to identify participants' joy of working on the contribution task relative to consuming leisure. We do so by controlling for monetary opportunity cost of work, i.e., by incorporating a paid leisure task. This novel technique allows us to make clear predictions on how much output a money-maximizing worker produces and to measure intrinsic motivation as deviations from this benchmark.[2] To the best of our knowledge, there is only one other paper analyzing how motivation moderates willingness to contribute to a public good. Tonin and Vlassopoulos (2015) find that participants who produce less output in a real effort task under a piece rate, increase output more compared to highly productive participants, when part of their earnings are donated to a charity. While their findings are in line with ours, we add by using a design that allows us to control for differences in productivity and therefore provides us with a cleaner measure of motivation.

2 An alternative to predict money-maximizing output is to exogenously vary effort cost as suggested by Gächter, Huang and Sefton (2016) which we discuss in the experimental design section of this paper. Furthermore, Della Vigna and Pope (2016) measure real effort cost by comparing effort between subjects under various payment schemes.

2 Experimental design

We first provide a general description of the main treatments before explaining the detailed parameters and procedures. The core element of our experiment is a simple but tedious work task in which workers are paid a piece rate to calculate cross sums for a pre-specified amount of time.

Table 1: Main treatments and phases

Treatment	**Phase 0** (5 minutes)	**Phase 1** (20 minutes)	**Phase 2** (20 minutes)
Individual (N = 51)	Player i gets 0.7 points per task completed by i	Same as Phase 0 **&** Opportunity cost of work (1 point per 15'' of leisure)	Same as Phase 1
Info (N = 51)	Player i gets 0.7 points per task completed by i	Same as Phase 1 in Individual **&** Assignment to group of 3, Info about absolute and relative output within group	Same as Phase 1
Team (N = 51)	Player i gets 0.7 points per task completed by i	Same as Phase 1 in Info	Same as Phase 2 in Info **&** Externality (players j and k get 0.2 per task completed by i, and vice versa)

Notes: The experiment had several non-incentivized elements before and after the main phases which are described in the main text. There were also a number of additional control treatments described in the appendix A 2.

Table 1 shows the three phases in the three main treatments we use to investigate the effect of observation and team incentives (there were in addition various unpaid elements like questionnaires preceding and following these phases, which will be explained later). Phase 0 is the same in all three treatments. In this phase, workers can work on the task and calculate cross sums from single digit numbers during 5 minutes. Cross sums are easy at first (e.g. 5 + 7 = ?) and get increasingly difficult over time as more and more digits are added to a cross sum (e.g. 8 + 9 + 4 + 2 + 1 + 5 + 7 + 8 + 9 + 1 = ?). Workers are paid a piece rate 0.7 points (which are converted into money at the end of the experiment) per correct cross sum. This means that workers take more and more time per task since they calculate the first few cross-sums quickly and then slow down. The purpose of phase 0 is to obtain a measure of workers' productivity. In this phase, workers have incentives to calculate as many cross sums as they can.

Phase 1 adds opportunity cost of work in the guise of a paid "leisure task" in all main treatments. To switch from the work task to the leisure task, a worker can press a button at any time during the 20 minutes of phase 1, in which case a screen with a countdown appears, adding 1 point to his or her earnings every 15 seconds. Through the leisure task, we induce interior optimal, i.e., money maximizing, behavior depending on productivity for each participant separately. Intuitively, a money-maximizing worker should stop working if it takes him or her longer than 15 seconds to earn a point by calculating (see below for detailed explanations). This money-maximizing behavior is constant over all treatments and therefore can be used as a reference point for the analysis of team effects. In addition, it is a key element of our measurement of motivation, which we use as moderator of the effect of observation and team incentives. To be more precise, a worker enjoying doing the task compared to sitting in front of an empty screen (= leisure) may work longer than money maximizing, a worker having a disutility from doing the task may stop early.

We are not the first to provide workers with opportunity cost of work by offering a paid alternative to effort provision (e.g., Berger, Harbring and Sliwka 2013, Eckartz 2014, Erkal et al. 2018, Mohnen et al. 2008, Schram and Weber 2017).[3] However, we are, to the best of our knowledge, the first to use a task with induced monetary opportunity cost of work to study optimal labor supply and to measure motivation to work. A close match to our experimental design is Gächter et al. (2016). They have developed a "ball-catching task" for which the experimenter can vary the cost of effort exogenously. Therefore, they can, as we do, estimate point predictions of money-maximizing behavior for their participants. However, while differences in productivity are induced by different cost functions defined by the experimenter in their setting, we estimate money-maximizing output for participants. While their approach has the advantage of tight control, ours has the advantage of allowing for heterogeneity in productivity which is relevant because productivity might be correlated with motivation. In addition, we discuss possible causes of deviations from money-maximizing behavior, in particular motivation, and show how they interact with incentives.

In the treatments Info and Team, there is an additional change from phase 0 to 1. In both of these treatments, workers are informed at the beginning of phase 1 that they are now in groups of three, and that all group members will be informed at the end of phase 1 about the relative and absolute performance of all group members (i.e., how many cross sums each has completed). Providing that information, we can measure the effect of (anticipated) observation on work effort by comparing output in phase 1 in Individual and Info.

Phase 2 brings no changes in treatment Individual. The comparison of phases 1 and 2 therefore serves to isolate potential learning and fatigue effects. As participants in Info receive information on relative performance between phases 1 and 2, we can compare output in phase 1 in Individual and Info as well as the change in output between phases 1 and 2 (diff-in-diff) in these treatments to evaluate the effect of observation. In treatment Team, additionally to the information on relative performance, we introduce team incentives in phase 2. Each group member's output exerts a positive externality on the

[3] In addition there are also several papers offering unpaid alternatives as for example Abeler et al. (2011), Charness, Masclet and Villeval (2014), Corgnet et al. (2015b), or Kessler and Norton (2016), or the option to leave the lab (see Dickinson 1999 or Rosaz, Slonim and Villeval 2016). We cannot use either of these alternatives as we need monetary opportunity cost in order to predict money-maximizing behavior.

other two group members. In particular, players *j* and *k* get 0.2 points per task completed by *i*, and vice versa. We again use a diff-in-diff approach to measure the effect of team incentives on output in a setting with observation: The difference of phases 1 and 2 in Team vs. the difference of phases 1 and 2 in Info. The combined effect of observation and team incentives is measured by comparing the change in performance between phases 1 and 2 in Individual and Team.

Procedures and parameters. Experiments were conducted in the laboratory of the Vienna Center for Experimental Economics using z-Tree (Fischbacher 2007). A total of 225 undergraduate students were recruited using ORSEE (Greiner 2015), of which 153 workers participated in the main treatments. Almost half of the workers were female (45%) and they work on average 8.56 hours per week.[4]

Upon arriving in the lab, all workers take a test for cognitive ability (20 minutes version of Raven's Advanced Progressive Matrices, see Hamel and Schmittmann 2006). The test scores are used as a control for low cognitive ability and lack of understanding of instructions as a potential cause of deviation from the benchmark for a money maximizer. After the test for cognitive ability, workers read the instructions (see appendix C) and are provided four minutes of unpaid practice time with the work task. Participants are told that the task has a steep learning curve and that they can increase their earnings in the rest of the experiment if they practice the task (all do). After the practice phase, the three main phases of incentivized real-effort tasks follow as shown in *Table 1*. Points are converted into money at the exchange rate of 15 points = 1 Euro and paid out in private at the end of the experiment. After the main phases, workers answer a short questionnaire, mainly on socio-demographics. On average, a session takes 90 minutes and workers earn €22.

The work task we use is simple but tedious and is similar to tasks used in other real-effort experiments (e.g., Bartling et al. 2009, Corgnet et al. 2015b, Dohmen and Falk 2011, Eriksson et al. 2009, Niederle and Vesterlund 2007). In particular, workers are shown three cross sums consisting of single digit numbers per screen (see appendix C for screenshots). The cross sums first consist of only two numbers and become longer, i.e., more tedious to complete, over time. Every five screens, one digit is added to the cross sum. This means, for example, that as participants have completed 105 tasks, they face cross sums with 9 digits for the next five screens (or 15 tasks). Participants receive a piece rate of 0.7 points per task, or, as each screen shows three tasks, 2.1 points per screen. When workers have calculated the three cross sums, they need to press an "ok" button to move to the next screen. The workers can only move on to the next screen if they have calculated all the cross sums shown on the screen correctly. If a worker makes one or more mistakes, an error message appears (saying for example "Answer 1 is wrong.") and the worker has to correct the mistake.

[4] Characteristics of participants in the control treatments are very similar.

The paid leisure task allows us to calculate a benchmark for the optimal switching point for a money maximizer. This money-maximizing switching point is reached when it takes a worker longer to earn a point by working than by looking at the countdown on the "leisure task" screen. Given our parameters, this is the case if it takes a worker more than 31.5 seconds (= 2.1 points per screen / 1 point every 15 seconds in leisure) to complete the three tasks on a screen.[5] This information is also provided to the participants und we ask control questions before phase 1 starts, to control for workers' understanding of this switching point (see appendix C for instructions).

In phase 2 of treatment Team, workers are still paid a piece rate of 0.7 points, but now also each team member receives 0.2 points per task. This incentive scheme does not change the benchmark for a money maximizer, but adds a social aspect to effort provision. As the team earns 0.7 + 2 * 0.2 = 1.1 points per task, a group-minded worker, i.e., someone trying to maximize group return, should stop working on the real-effort task when it takes him or her more than 49.5 seconds (= 3.3 points per screen / 1 point every 15 seconds in leisure) to complete the three tasks on a screen. Thereby, the Team treatment is isomorphic to a public good game with an interior Nash equilibrium (the benchmark for a money maximizer) and an interior social optimum (optimal behavior for the group-return maximizer).[6] Note that it is not socially optimal to spend the entire time working (for someone working at a reasonable speed) because of the increasing difficulty of the task. Compared to previous studies, our task has the great advantage of a known interior solution, which also means that workers have scope to respond to incentives.[7]

3 Estimating optimal output and intrinsic motivation

This section models how much a worker in a team wants to work. We assume the worker's output depends on his or her revenues, on how much he or she likes the job (i.e., whether he or she is intrinsically motivated), on how much he or she cares about coworkers who benefit from his or her effort, and the extent to which he or she is susceptible to peer pressure (i.e., to being observed). For

[5] We refer to the time per screen here because our software only provides us with a measure (a time stamp) when subjects click the ok-button to move to the next screen and not when typing in a cross sum.

[6] Literature has shown that giving in public good games is very robust and is thus a good way to add cooperation incentives, without changing individual monetary incentives (Andreoni 1995).

[7] Many previous studies failed to find positive labor supply responses to increasing wages because they implemented corner solutions in which it was money maximizing for subjects to work all the time (as in our phase 0) (Cappelen et al. 2013, Corgnet et al. 2015b, Eckartz 2014).

simplicity, we assume zero effort cost of work.[8] We use the model to discuss interaction effects between these determinants with a special emphasis on the relation between intrinsic motivation and team incentives. Finally, we explain how we use the model to empirically identify intrinsic motivation.

Teamwork. Our basic assumptions are that the utility of worker *i* comes from working and from consuming leisure. The return for a money maximizer from working is the number of completed tasks Q_i multiplied with the piece rate w. Work is costly because workers forego leisure. Workers value leisure at *l* Euros per second.

Teamwork involves a positive externality, which we call team incentives. In the Team treatment, each task completed by worker *i* also benefits the other two team members *j* and *k* by increasing their payoff by *s*. If worker *i* has pro-social preferences, i.e., is cooperative, doing so increases *i*'s utility by γ_i. Additionally, each task completed by worker *i* increases *i*'s relative position in the team. We assume worker *i* receives a utility of ρ_i from increasing his or her position relative to the average of the other workers $\bar{Q}_{-i}$. Furthermore, participants receive a payoff *s* per task completed by their team members Q_{-i}. Therefore, a worker with prosocial preferences ($\gamma_i > 0$) and sensitivity for peer pressure ($\rho_i > 0$) has the following utility function:

$$U_{i_{\gamma>0,\rho>0}} = (w + \gamma_i)Q_i + l\big(E - T_i(Q_i)\big) + s \sum Q_{-i} + \rho_i(Q_i - \bar{Q}_{-i}). \tag{1}$$

A key innovation of our design is that we introduce a *paid* "leisure task" which means that leisure has a controlled (i.e., experimentally induced) monetary value. The total time available for work or leisure is *E*, and T_i is the total time worker *i* spent working, i.e., $E - T_i$ is the time spent in leisure. The money-maximizing number of tasks depends on how much time it takes a worker to complete a task (i.e., on his or her productivity).

The condition for the optimal switch point *for a money-maximizing worker*, i.e., $\rho_i = 0$ & $\gamma_i = 0$ is:

$$\frac{\partial T_i}{\partial Q_i} = \frac{w}{l} \ . \tag{2}$$

Eq. (2) shows that a money-maximizing worker should stop working when the time needed to complete a task exceeds the ratio of marginal revenue to marginal cost. To derive a closed expression

[8] However, it is straightforward to present the model including (possibly non-linear) effort cost. Implications of such costs are discussed at the end of this section.

for the optimal output we assume a quadratic *cost function* as shown in eq. (3), with β_{1i} and β_{2i} representing productivity parameters which we estimate: [9]

$$T_i(Q_i) = \beta_{1i}Q_i + \beta_{2i}Q_i^2\,. \tag{3}$$

Maximizing eq. (1) w.r.t. to Q_i using the cost function defined in eq. (3) leads to the optimal output for a worker with pro-social preferences (γ_i > 0) and sensitivity for peer pressure (ρ_i > 0) of:

$$Q^*_{i_{\gamma_i>0,\rho_i>0}} = \frac{1}{2\beta_{2i}}\left(\frac{w+\rho_i+\gamma_i}{l} - \beta_{1i}\right). \tag{4}$$

If a worker is a pure money maximizer, i.e., has neither pro-social preferences nor sensitivity to peer pressure ($\gamma_i = 0, \rho_i = 0$), or is working in the Individual treatment in which there are no teams, the optimal output reduces to:

$$Q^*_{i_{\gamma_i=0,\rho_i=0}} = \frac{1}{2\beta_{2i}}\left(\frac{w}{l} - \beta_{1i}\right). \tag{5}$$

Intrinsic motivation. As a next step, we add intrinsic motivation μ_i to the utility function in eq. (1). A positive μ_i means that a worker prefers working compared to leisure and vice versa for a negative μ_i. Note that this measure is not an absolute characteristic of the worker, but it reflects a person's utility of working in a particular task compared to a particular type of leisure. This relative measure in line with the standard measure of intrinsic motivation in the literature, namely effort provided in absence of performance related incentives (Gneezy and Rustichini 2000). For simplicity, we use the same function (a log-function) to model positive and negative intrinsic motivation:

$$U_{i_{\gamma_i>0,\rho_i>0,\mu_i\neq 0}} = (w+\gamma_i)Q_i + l\left(E - T_i(Q_i)\right) + s\sum Q_{-i} + \rho_i(Q_i - \bar{Q}_{-i}) + \mu_i\, log(Q_i). \tag{6}$$

Positive intrinsic motivation shifts the utility function up, i.e., for each task completed the utility is higher with positive intrinsic motivation, compared to a money-maximizer. However, the marginal benefit from work falls with Q for a worker with positive intrinsic motivation, and the marginal cost decreases for a worker with negative intrinsic motivation. These assumptions are in line with the idea that both motivations wash out with the activity. The optimal output for a prosocial worker, i.e., for a worker with social preferences γ_i > 0 and a preference for relative standing ρ_i> 0, and intrinsic motivation $\mu_i \neq 0$ is:

[9] $T_i(Q_i)$ shows how much time it takes to produce output (i.e., to calculate the cross sums). This function can be thought of as a cost function because time spent working cannot be used on leisure, which, in turn, has an induced monetary opportunity cost. Appendix A1 shows that quadratic cost functions approximate the observed data very well.

$$Q^*_{i_{\gamma_i>0,\rho_i>0,\mu_i\neq 0}} = \frac{w+\rho_i+\gamma_i-\beta_{1i}l+\sqrt{(w+\rho_i+\gamma_i-\beta_{1i}l)^2+8l\mu_i\beta_{2i}}}{4l\beta_{2i}} . \tag{7}$$

Interaction effects. The model allows for the possibility of heterogeneous effects, i.e., that the response to team incentives differs across workers' preferences, including their intrinsic motivation, and their productivity. Heterogeneous effects are captured by cross-partial derivatives. As the optimal output depends on productivity represented by β_{1i} and β_{2i}, we take the cross-partial derivative of eq. (4) w.r.t. γ_i (the same holds for ρ_i) and the productivity parameters β_{1i} and β_{2i} to analyze whether heterogeneous workers react with a different intensity to team incentives. While the cross-partial derivative w.r.t. β_{1i} is 0, eq. (8) shows that the effect of team incentives is stronger for productive workers, i.e. those who are fast at completing hard tasks (a low β_{2i} indicates a more able participant):

$$\frac{\partial^2 Q^*_{i_{\gamma_i>0,\rho_i>0}}}{\partial\gamma_i\partial\beta_{2i}} = -\frac{1}{2\beta_{2i}^2 l} < 0 . \tag{8}$$

Eq. (9) shows that the marginal effect of team incentives on optimal output decreases in intrinsic motivation μ_i:

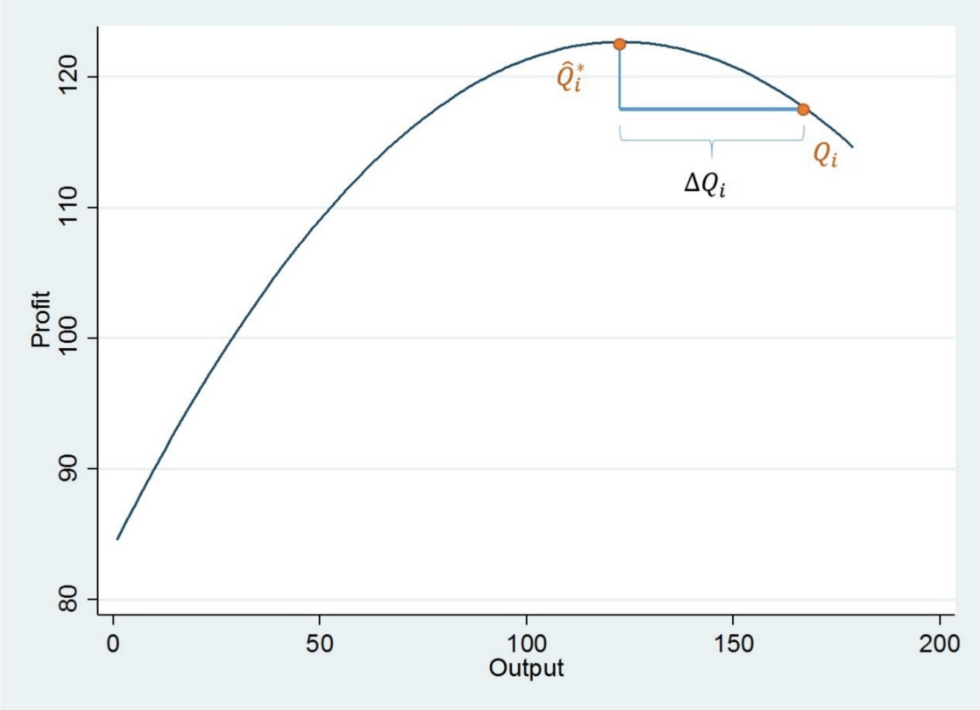
$$\frac{\partial^2 Q^*_{i_{\gamma_i>0,\rho_i>0,\mu_i\neq 0}}}{\partial\gamma_i\partial\mu_i} = -\frac{(w+\rho_i+\gamma_i-\beta_{1i}l)}{((w+\rho_i+\gamma_i-\beta_{1i}l)^2+8l\mu_i\beta_{2i})^{\frac{3}{2}}} < 0. \tag{9}$$

Intuitively, this means that if a participant is positively intrinsically motivated and thereby works more beyond what is money-maximizing, he or she will increase output less in the Team treatment than a participant who has produced output below what is money maximizing, i.e., was negatively motivated.

Measuring intrinsic motivation in the experiment. Preferences cannot be directly observed. This also applies to intrinsic motivation. We infer intrinsic motivation from the deviation of observed output choices from optimal (counterfactual) output Q^* for a money-maximizing worker, i.e. a worker who has no intrinsic or social motivation. That is, $\Delta Q_i = Q_i - \hat{Q}_i^*$ is our measure of intrinsic motivation (see *Figure 1*).

How do we estimate the counterfactual $\hat{Q}_i^*$? Conceptually, Q^* depends on the productivity of a worker, which, unfortunately, cannot be directly observed. However, we can observe how long it takes a worker to complete the tasks and we can estimate a production (cost) function from these observations. In particular, we use time used per task and the number of completed tasks to estimate eq. (3) for each participant to obtain $\hat{\beta}_{1it}$ and $\hat{\beta}_{2it}$. These estimates are used to obtain $\hat{Q}_i^*$ for each participant using eq. (5) (see appendix A 1 for details).

Figure 1: Measuring intrinsic motivation



Notes: $\hat{Q}^*_{i_{\gamma_i=0,\rho_i=0,\mu_i=0}}$ is the benchmark output (number of completed work tasks) of a money maximizer and Q_i is the observed number of completed tasks for a positively motivated worker ($\mu > 0$).

Identifying intrinsic motivation. The optimal output of a money maximizer provides the benchmark to measure intrinsic motivation to work hard, $\Delta Q_i = Q_i - \hat{Q}_i^*$. To cleanly identify intrinsic motivation deviations from optimal output need to be systematic and deliberate rather than driven by error or bounded rationality. To control for these factors as far as possible, we saliently inform participants in our experiment when it is optimal for a money maximizer to switch to leisure (see instructions in appendix C). Evidence suggests that deviations from money-maximizing output are unlikely to be caused by bounded rationality because the deviations are not correlated with a proxy of cognitive ability (Pearson r = -0.14, p = 0.326, for ΔQ_1 & r = -0.20, p = 0.158, for ΔQ_2, see also *Figure A3*). We can reject the hypothesis that output choices are random because individual intrinsic motivation ΔQ_{i1} and ΔQ_{i2} is highly correlated across phases (Pearson r = 0.62, p = 0.000). This suggests that individual deviations from the money-maximizing benchmark are systematic.

Recall that we define intrinsic motivation as how enjoyable or boring it is to work relative to consuming leisure, net of effort costs. If participants experience such effort cost, the observed $\Delta \mathrm{Q}_i$ underestimates true intrinsic motivation. While there certainly are environments in which non-monetary effort cost are large (think for example about physical work), we expect non-monetary effort cost to be negligible given the choice of task and the limited duration of the experiment. This expectation is supported by empirical findings showing that the elasticity of effort with respect to incentives is low in real-effort laboratory experiments unless there are monetary cost of work (see for example Araujo et al. 2016, Eckartz 2014 and Gächter et al. 2016 for a discussion) indicating that cognitive effort cost are small. Furthermore, we do not find a positive correlation between productivity (measured as the time per screen for the first ten screens) and ΔQ_i (Pearson *r* = 0.01, *p* = 0.898, *N* = 147), indicating that there is no systematic difference in intrinsic motivation depending on productivity.[10] Another possible confound is that motivation affects the speed at which workers compute the tasks, because working fast is more costly than working slow. This endogeneity would lead to an underestimation of intrinsic motivation for highly productive participants, as we would overestimate their money-maximizing benchmark. We run an additional control treatment in which we compare calculation speed (measured as the time per screen for the first ten screens) under different piece rates and do not find any differences (*p* = 0.697, two-sided *t*-test). We conclude that the difference in effort cost for different calculation speeds, if any, are dominated by monetary incentives and therefore negligible. In all, we conclude that ΔQ_i is a valid measure of intrinsic motivation.[11]

4 Results

Section 4.1 describes money-maximizing output and intrinsic motivation in general, section 4.2 discusses treatment effects and section 4.3 shows that personal characteristics, i.e., differences in productivity and motivation, shape the effect of team incentives and observation on output.

Randomization of workers across treatments worked well. For example, the share of females, the average work hours per week, and the social structure are not statistically different across treatments.[12] However, average payoffs are different (all treatment comparisons are significant at the 5%-level or better) due to the externality.

4.1 Intrinsic motivation and money-maximizing output

We find that workers are on average motivated to work on the job, and this motivation is constant across phases. In fact, average intrinsic motivation ΔQ is positive, i.e., they work more than is money maximizing (*p* = 0.001, *t*-test), and ΔQ is constant across all treatments in phase 1 (Individual vs. Info: 7.85 vs. 8.83 tasks, *p* = 0.861; Individual vs. Team: 7.85 vs. 6.22 tasks, *p* = 0.719, Wald test, based on

[10] If the cognitive effort cost are increasing in output (for example due to the increasing task difficulty), underestimation of intrinsic motivation is more pronounced for highly productive participants, leading to a biased estimate of intrinsic motivation.

[11] Appendix A 2 discusses possible biases in our estimates and provides robustness checks.

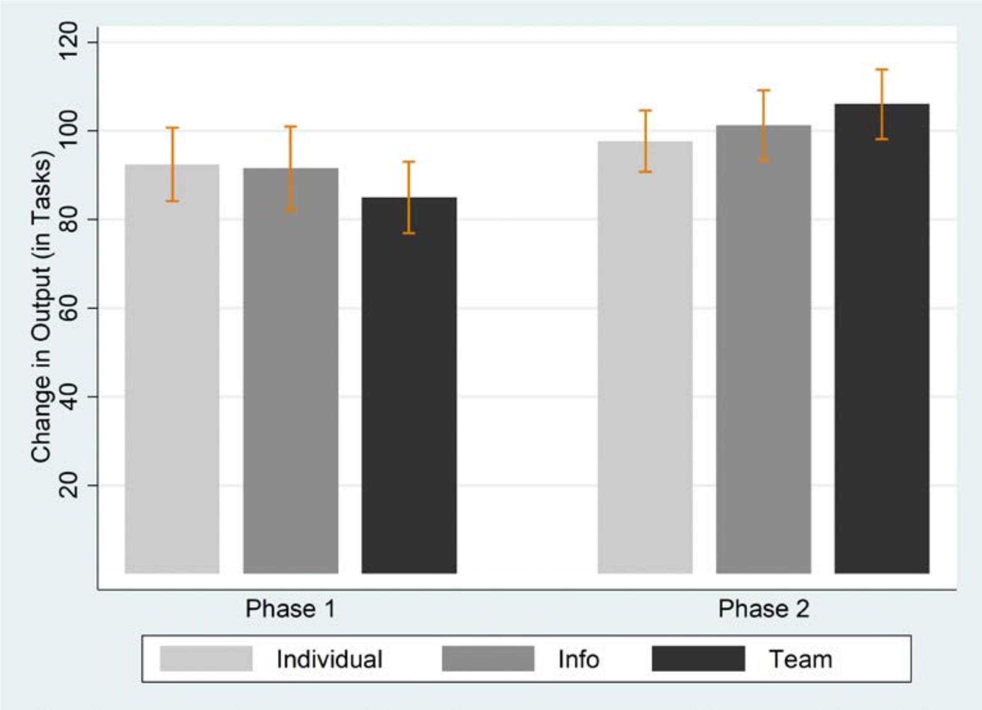
[12] The share of females is in Individual 35% vs. Info 57% vs. Team 43%, p = 0.094, Fisher's exact test. The average work hours per week are in Individual: 6.39 vs. Info: 8.90, p = 0.228 & Info: 8.90 vs. Team: 10.37, p = 0.534, two- sided t-tests. The social structure is upper middle class 37.25% vs. 39.22% vs. 43.14%, lower middle class 33.33% vs. 31.37 vs. 29.41, working class 27.45% vs.29.41% vs. 27.45%, upper class 1.96% vs. 0% vs. 0%, p = 0.987, Fisher's exact test.

an OLS regression with standard errors clustered by teams). This absence of differences is not surprising because incentives are the same in phase 1 in all treatments. Incentives are constant across phases in the Individual treatment. Behavioral differences across phases in this treatment, if any, show learning and fatigue effects. It is important to check for the presence of such effects because they potentially confound our measure of intrinsic motivation. Fortunately, there are no such effects (ΔQ is constant, p = 0.709, paired t-test).

4.2 Team incentives and observation

We find that observation of one's team members' performance by itself had no effect but coupled with team incentives, a significant effect on average output obtains.

Figure 2: Team incentives and observation increase output



Notes: Bars show mean output in phases 1 and 2 for treatments Individual (light grey bars), Info (dark grey bars) and Team (black bars). Whiskers show 95% confidence intervals. We have 51 observations per treatment. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.

Figure 2 shows a significant increase in output in the Team treatment of 25% (from 86 to 107 tasks, compare black bars, whiskers show 95% confidence intervals). This effect is significantly bigger than the increase in the Individual treatment of 3% (from 93.96 to 96.72, compare light grey bars).[13] That is, the difference-in-difference comparison is significant (p = 0.000, Wald test). The difference-

[13] There is no significant difference in output between treatments in phase 1 (Individual vs. Info: p = 0.822, Individual vs. Team: p = 0.153 and Info vs. Team: p = 0.102, Wald tests). *Table B1* provides detailed complementary OLS regressions.

in-difference test for team incentives alone is also significant: output in Team increases by 13.90 tasks more than in Info (p = 0.001, Wald test). Taken together, this evidence indicates that the introduction of team incentives but not the information provided drives the increase in average output.

The absence of observation effects on average masks heterogeneous effects with respect to gender. While men do not react to feedback received between phases (Individual vs. Info: +5.81 tasks vs. +3.30 tasks, p = 0.470, Wald tests), women in Info increase output significantly across phases compared to women in the Individual treatment (Individual vs. Info: -2.67 vs. 10.03, p = 0.054, Wald tests).

4.3 Motivation shapes the effectiveness of team incentives and observation

The previous section has shown that the Team treatment has an effect overall, i.e., output increases more in Team than in Individual. We now discuss heterogeneous treatment effects which are predicted from our simple model in section 3. In particular, eq. (9) predicts that those with low intrinsic motivation react more strongly to team incentives and observation than those with high intrinsic motivation. In line with these predictions, we find that workers with high intrinsic motivation do not react to the Team treatment, but those with low intrinsic motivation and money maximizers do react.

Figure 3 shows the change in output from phase 1 to 2 for different levels of intrinsic motivation, by treatment. Positive motivation means that workers complete more tasks than is money maximizing, and vice versa for negative motivation. In particular, we split participants in each treatment into terciles based on motivation in phase 1 (see Table B2 and Table B3).14 We find evidence for learning effects in the sense that workers in Individual tend to move closer to $\hat{Q}_i^*$ in phase 2. That is, those in the negative motivation tercile (see leftmost light grey bar) tend to increase and those in the positive tercile (rightmost light grey bar) tend to decrease output (+ 10.06 tasks vs. -2.44 tasks, p = 0.051, Mann-Whitney test).

We use an OLS regression with interaction terms for motivation type to control for these learning effects and, hence, to show that motivation shapes the effectiveness of team incentives and observation (see *Table B4*). In Team, when workers observe each other and are subject to team incentives, negatively motivated workers increase output by 33.18 tasks (see leftmost black bar) while positively motivated workers increase output by a mere 1.68 tasks (see rightmost black bar). The difference in absolute increase is 31.50 tasks. That is, negatively motivated workers increase output about 20 times more than the positively motivated. This difference is significantly bigger than the learning effect in Individual

[14] The results below also hold for alternative definitions of positive and negative motivation (see online appendix).

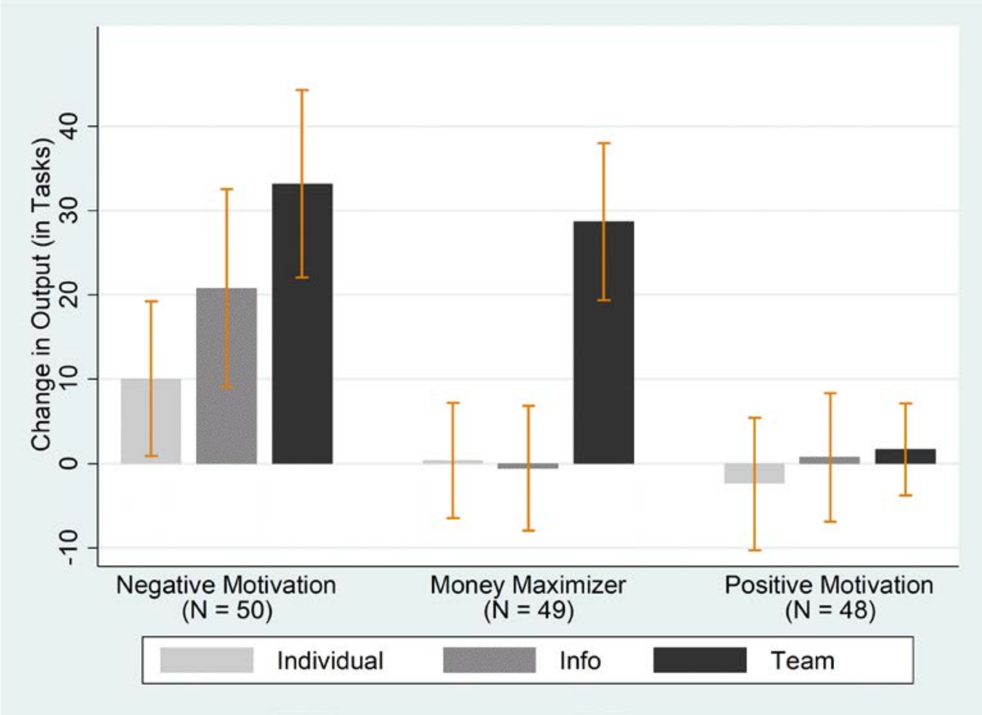
(12.50), $p = 0.001$. We conclude that negatively motivated workers respond more to team incentives and observation than positively motivated workers.

The same conclusion holds for money maximizers in the Team treatment. They increase output by 28.68 (see middle black bar) while positively motivated workers do so by only 1.68, for a difference of 27.00. In Individual, money maximizers increase output by 0.35 (see middle light grey bar), while positively motivated workers reduce output by -2.44, for a difference of 2.79. Hence, money maximizers increase output more than positively motivated workers, controlling for learning effects (Individual (2.79) vs. Team (27.00), $p = 0.000$).[15]

The effect of observation alone is not significantly shaped by intrinsic motivation. In Info, negatively motivated workers increase output by 20.82 tasks (see leftmost dark gray bars), while positively motivated workers increase output only by 0.75 tasks (see rightmost dark gray bars). This difference is not significantly bigger than the corresponding difference in Individual (Individual (12,50) vs. Info (20.07), $p = 0.373$). This lack of an effect of observation suggests that the interaction effect is mainly driven by team incentives. To test, we compare the difference in output change between positive vs. negative motivation in Team vs. Info. In Team, we find that the difference in the change in output between workers with negative (33.18) and workers with positive (1.68) motivation is 31.50. In Info, the difference in output between workers with negative (20.82) and workers with positive (0.75) motivation is 20.07. Therefore, the difference between the respective changes in Team vs. Info is 11.43 tasks ($p = 0.183$) and 28.31 tasks between money maximizers and workers with positive motivation (Info (-1.31) vs. Team (27.00) $p = 0.000$), respectively.

[15] Our result is in line with Chen et al. (2010) who also consider a setting with observation (information on relative performance) and team incentives (positive externalities from contributions to an online community). They find that observation and team incentives increase effort for participants with a lower ranking. The low ranking may in principle be due to motivation or ability. Since these authors analyze performance in rating movies, the ranking is likely driven mainly by motivation rather than ability.

Figure 3: Motivation shapes the effectiveness of team incentives and observation



Notes: Bars show mean change in output between phases for different motivation terciles. Whiskers show 95% confidence intervals. Splitting the subjects into terciles results in bins of 51 observations each. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.

Is the effect of team incentives and observation also shaped by productivity? Our model from section 3 predicts that more able workers react more strongly to team incentives as well as observation (see eq. 8). In line with these predictions we find that when productivity increases by one standard deviation, the combined effect of team incentives and observation on output increases by 8.91 tasks (p = 0.023, see OLS regression provided in *Table B5*).[16] This result is mainly driven by productivity shaping the effect of observation. In fact, an increase of one standard deviation in productivity increases the effect of observation by 8.17 tasks (p = 0.061, Wald test).

To sum up, we find that output in the Team treatment increases significantly less for positively motivated workers than for negatively motivated workers and money maximizers, controlling for learning effects. Furthermore, we find stronger treatment effects for more productive workers.

Why does motivation shape the effectiveness of team incentives and observation? One candidate explanation is reciprocity concerns. Workers in Info and Team observe their relative performance, i.e. their rank. Workers with rank 3 work less than workers with rank 1. Therefore, those with rank 3 receive more from other players than those with rank 1 in the Team treatment. Consequently, reciprocal workers

[16] We find similar results when using the time per screen as an alternative productivity measure (see *Table B6*).

with rank 3 increase their output more than reciprocal workers with rank 1. Positive motivation increases the chance to end up in the highest rank (see *Table B8*). Therefore, reciprocity concerns could explain our results for motivation. To test, we investigate if the effectiveness of team incentives and observation depends on ranking. We find no such effects and conclude that reciprocity does not drive our results (see appendix B3).

We think our findings are best explained by the following simple cost consideration. Expanding output is profitable for those with output below the money-maximizing benchmark (i.e., those with negative intrinsic motivation) but costly for those beyond the money-maximizing benchmark (i.e., those with positive intrinsic motivation). Positive externalities in the Team treatment provide incentives to expand output for cooperative workers. But such expansion is more costly to those with positive motivation than those with negative motivation. Hence, these cost considerations imply that those with negative intrinsic motivation react more strongly to team incentives than those with positive intrinsic motivation.

5 Conclusion

This paper measures intrinsic motivation to work hard in a real-effort experiment and shows that intrinsic motivation importantly shapes the incentive effects of teamwork. We find that those with positive intrinsic motivation, i.e., those who like working on the task, respond less to team incentives and observation by peers than those who do not like the job. We demonstrate these effects in a novel experimental design, which allows us to estimate the optimal output for a money-maximizing worker by controlling monetary opportunity cost of work. We use the resulting estimate as a benchmark to measure intrinsic motivation. This measure is clean as it excludes other motivators to work hard, such as image concerns or social ties at the workplace. We then use this measure to show that intrinsic motivation shapes the effectiveness of team incentives in the guise of positive spillovers between workers when effort is observable. In particular, we find that spillovers have a large effect when workers have low intrinsic motivation but a small effect when workers have high intrinsic motivation.

We believe our contribution is relevant both for experimental research and for managerial practice. First, our novel experimental design with controlled opportunity cost of work can be used to measure biases (like overconfidence) and individual responses to incentives (like the disincentive effects of redistribution, Sausgruber et al. 2018) in the laboratory.

Second, our key substantive finding may be relevant for managerial practice. The finding is that motivation shapes the effectiveness of team incentives and observation by peers, and we think it is relevant because it is consistent with a simple and general cost argument. The cost argument can be applied both across workers and across jobs. To illustrate how different types of workers respond to team incentives, think of two workers. She is highly intrinsically motivated, works twelve hours per day, while he is not, and only works eight hours per day. Who is more likely to increase working hours if they have to work in a team? As the highly motivated worker already works long hours, she is unlikely to increase her work time when put in a team, as doing so is more costly to her than for him who works short hours. The reason is that she already has less leisure than he does, and the relative value of leisure increases with its scarcity. This cost argument also applies to predictions across different types of jobs. Workers are more motivated to work hard on jobs that are intrinsically rewarding (Reynolds and Aletraris 2007). Our results suggest that workers in interesting and challenging jobs react less strongly to team incentives than those who work on a meaningless and repetitive task. In conclusion, because the cost argument applies both across workers and jobs, our results suggest that managers who have to choose whom to assign to team tasks vs. individual tasks, should assign the non-intrinsically motivated workers and those who work in strenuous jobs to teams.

Third, whether our approach can be used to classify workers as positively vs. negatively motivated in managerial practice must remain an open question. The reason is that our classification of workers as positively motivated is based on the performance in a specific environment and task (calculating cross-sums), but a given worker may be classified differently if working in a different setting. Indeed, our measure may capture a preference for calculating cross-sums rather than a more general preference for working. However, Kanfer, Frese and Johnson (2017) show that the motivation to work hard results from a combination of individual traits (like conscientiousness), the specific job characteristics (like strenuousness), and the work environment (like the organizational culture). Testing the prognostic value of our measure of motivation along these three dimensions is interesting but beyond the scope of this paper. Instead, we focus on the important issue of how motivation shapes the effectiveness of team incentives and observation.

References

Abeler, J., Falk, A., Goette, L., Huffman, D. 2011. Reference Points and Effort Provision. *American Economic Review*. **101**(2) 470-492.

Andreoni, J. 1995. Cooperation in Public-Goods Experiments - Kindness or Confusion. *American Economic Review*. **85**(4) 891-904.

Araujo, F.A., Carbone, E., Conell-Price, L., Dunietz, M., Jaroszewicz, A., Landsman, R., . . . Wilson, A. 2016. The Slider Task: An Example of Restricted Inference on Incentive Effects. *Journal of the Economic Science Association*. **2**(1) 1-12.

Azmat, G., Iriberri, N. 2016. The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction. *Journal of Economics & Management Strategy*. **25**(1) 77-110.

Babcock, P., Bedard, K., Charness, G., Hartman, J., Royer, H. 2015. Letting Down the Team? Social Effects of Team Incentives. *Journal of the European Economic Association*. **13**(5) 841-870.

Balafoutas, L., Kocher, M.G., Putterman, L., Sutter, M. 2013. Equality, Equity and Incentives: An Experiment. *European Economic Review*. **60** 32-51.

Bandiera, O., Barankay, I., Rasul, I. 2010. Social Incentives in the Workplace. *Review of Economic Studies*. **77**(2) 417-458.

Bandiera, O., Barankay, I., Rasul, I. 2013. Team Incentives: Evidence from a Firm Level Experiment. *Journal of the European Economic Association*. **11**(5) 1079-1114.

Barankay, I. 2012. Rank Incentives: Evidence from a Randomized Workplace Experiment. *Working Paper*. Wharton School, University of Pennsylvania

Bartling, B., Fehr, E., Marechal, M.A., Schunk, D. 2009. Egalitarianism and Competitiveness. *American Economic Review*. **99**(2) 93-98.

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.

Berger, J., Harbring, C., Sliwka, D. 2013. Performance Appraisals and the Impact of Forced Distribution-an Experimental Investigation. *Management Science*. **59**(1) 54-68.

Berger, J., Herbertz, C., Sliwka, D. 2011. Incentives and Cooperation in Firms Field Evidence. *IZA Discussion paper series*

Bognanno, M., Huffman, D. 2017. High-Powered Performance Pay and Crowding out of Non-Monetary Motives. *Management Science*. Forthcoming

Boning, B., Ichniowski, C., Shaw, K. 2007. Opportunity Counts: Teams and the Effectiveness of Production Incentives. *Journal of Labor Economics*. **25**(4) 613-650.

Bowles, S., Polania-Reyes, S. 2012. Economic Incentives and Social Preferences: Substitutes or Complements? *Journal of Economic Literature*. **50**(2) 368-425.

Budescu, D.V. 1993. Dominance Analysis - a New Approach to the Problem of Relative Importance of Predictors in Multiple-Regression. *Psychological Bulletin*. **114**(3) 542-551.

Cappelen, A.W., Moene, K.O., Sorensen, E.O., Tungodden, B. 2013. Needs Versus Entitlements an International Fairness Experiment. *Journal of the European Economic Association*. **11**(3) 574-598.

Charness, G., Masclet, D., Villeval, M.C. 2014. The Dark Side of Competition for Status. *Management Science*. **60**(1) 38-55.

Chen, Y., Harper, F.M., Konstan, J., Li, S.X. 2010. Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens. *American Economic Review*. **100**(4) 1358,1398.

Cherry, T.L., Kroll, S., Shogren, J.F. 2005. The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab. *Journal of Economic Behavior and Organization*. **57**(3) 357-365.

Cooper, D.J., Saral, K.J. 2013. Entrepreneurship and Team Participation: An Experimental Study. *European Economic Review*. **59** 126-140.

Corgnet, B., Hernan-Gonzalez, R., Schniter, E. 2015a. Peer Pressure and Moral Hazard in Teams: Experimental Evidence. *Review of Behavioral Economics*. **2**(4) 379-403.

Corgnet, B., Hernan-Gonzalez, R., Schniter, E. 2015b. Why Real Leisure Really Matters: Incentive Effects on Real Effort in the Laboratory. *Experimental Economics*. **18**(2) 284-301.

Deci, E.L. 1971. Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology*. **18**(1) 105-115.

Deci, E.L. 1975. *Intrinsic Motivation*. Plenum Pr., New York [u.a.].

Della Vigna, S., Pope, D. 2016. What Motivates Effort? Evidence and Expert Forecasts. *Review of Economic Studies*. Forthcoming.

Dickinson, D.L. 1999. An Experimental Examination of Labor Supply and Work Intensities. *Journal of Labor Economics*. **17**(4) 638-670.

Dohmen, T., Falk, A. 2011. Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*. **101**(2) 556-590.

Dutcher, E.G., Salmon, T.C., Saral, K.J. 2015. Is "Real" Effort More Real? *MPRA Paper*.

Eckartz, K.M. 2014. Task Enjoyment and Opportunity Costs in the Lab - the Effect of Financial Incentives on Performance in Real Effort Tasks. *Jena Economic Research Papers* **2014-005**.

Erev, I., Bornstein, G., Galili, R. 1993. Constructive Intergroup Competition as a Solution to the Free Rider Problem - a Field Experiment. *Journal of Experimental Social Psychology*. **29**(6) 463-478.

Eriksson, T., Poulsen, A., Villeval, M.C. 2009. Feedback and Incentives: Experimental Evidence. *Labour Economics*. **16**(6) 679-688.

Erkal, N., Gangadharan, L., Boon, H.K. 2018. Monetary and Non-Monetary Incentives in Real-Effort Tournaments. *European Economic Review*. **101** 528-545.

Falk, A., Ichino, A. 2006. Clean Evidence on Peer Effects. *Journal of Labor Economics*. **24**(1) 38-57.

Fehr, E., Gachter, S. 2000. Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*. **14**(3) 159-181.

Filiz-Ozbay, E., Ozbay, E.Y. 2014. Effect of an Audience in Public Goods Provision. *Experimental Economics*. **17**(2) 200-214.

Fischbacher, U. 2007. Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*. **10**(2) 171-178.

Frey, B.S., Jegen, R. 2001. Motivation Crowding Theory. *Journal of Economic Surveys*. **15**(5) 589-611.

Gächter, S., Huang, L.B., Sefton, M. 2016. Combining "Real Effort" with Induced Effort Costs: The Ball-Catching Task. *Experimental Economics*. **19**(4) 687-712.

Georganas, S., Tonin, M., Vlassopoulos, M. 2015. Peer Pressure and Productivity: The Role of Observing and Being Observed. *Journal of Economic Behavior and Organization*. **117** 223-232.

Gjedrem, W.G., Kvaløy, O. 2018. Smells Like Team Spirit: An Experiment on Relative Performance Feedback. *CESifo Working Paper*. Munich.

Gneezy, U., Meier, S., Rey-Biel, P. 2011. When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives*. **25**(4) 191-209.

Gneezy U., Rustichini A. 2000. Pay Enough or Don't Pay at All. *The Quarterly Journal of Economics.* **115**(3) 791-810.

Greiner, B. 2015. Subject Pool Recruitment Procedures: Organizing Experiments with Orsee. *Journal of the Economic Science Association*. **1**(1) 114-125.

Gromping, U. 2007. Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. *American Statistician*. **61**(2) 139-147.

Hamel, R., Schmittmann, V.D. 2006. The 20-Minute Version as a Predictor of the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*. **66**(6) 1039-1046.

Hamilton, B.H., Nickerson, J.A., Owan, H. 2003. Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy*. **111**(3) 465-497.

Herbst, D., Mas, A. 2015. Peer Effects on Worker Output in the Laboratory Generalize to the Field. *Science*. **350**(6260) 545-549.

Hoppe, E.I., Kusterer, D.J. 2011. Behavioral Biases and Cognitive Reflection. *Economics Letters*. **110**(2) 97-100.

Hossain, T., Li, K.K. 2014. Crowding out in the Labor Market: A Prosocial Setting Is Necessary. *Management Science*. **60**(5) 1148-1160.

Irlenbusch, B., Sliwka, D. 2005. Incentives, Decision Frames, and Motivation Crowding out an Experimental Investigation. *IZA Discussion paper series*

Kanfer, R., Chen, G., Pritchard, R.D. 2008. *Work Motivation : Past, Present, and Future*. Psychology Press, New York.

Kanfer, R., Frese, M., Johnson, R.E. 2017. Motivation Related to Work: A Century of Progress. *Journal of Applied Psychology*. **102**(3) 338-355.

Kessler, J.B., Norton, M.I. 2016. Tax Aversion in Labor Supply. *Journal of Economic Behavior and Organization*. **124** 15-28.

Luchman, J.N. 2013. *Domin: Stata Module to Conduct Dominance Analysis*. Boston College Department of Economics.

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

Mc Dowell, T., Agarwal, D., Miller, D., Okamoto, T., Page, T. 2016. *Organizational Design* **-** *the Rise of Teams*.

Mohnen, A., Pokorny, K., Sliwka, D. 2008. Transparency, Inequity Aversion, and the Dynamics of Peer Pressure in Teams: Theory and Evidence. *Journal of Labor Economics*. **26**(4) 693-720.

Muelbacher, S., Kirchler, E. 2009. Origin of Endowements in Public Good Games: The Impact of Effort on Contributions. *Journal of neuroscience, psychology, and economics*. **2**(1) 59-67.

Nalbantian, H.R., Schotter, A. 1997. Productivity under Group Incentives: An Experimental Study. *American Economic Review*. **87**(3) 314-341.

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

Reuben, E., Tyran, J.R. 2010. Everyone Is a Winner: Promoting Cooperation through All-Can-Win Intergroup Competition. *European Journal of Political Economy*. **26**(1) 25-35.

Reynolds, J., Aletraris, L. 2007. *For Love or Money? Extrinsic Rewards, Intrinsic Rewards, Work–Life Issues, and Hour Mismatches*. Workplace Temporalities.

Rosaz, J., Slonim, R., Villeval, M.C. 2016. Quitting and Peer Effects at Work. *Labour Economics*. **39** 55-67.

Sausgruber, R. 2009. A Note on Peer Effects between Teams. *Experimental Economics*. **12**(2) 193-201.

Sausgruber, R., Sonntag, A., Tyran, J.-R. 2018. Disincentives from Redistribution: Evidence on a Dividend of Democracy. *Working Paper*

Schram, A., Weber, M. 2017. The Non-Equivalence of Labour Market Taxes: A Real-Effort Experiment. *The Economic Journal*.

Segal, C. 2012. Working When No One Is Watching: Motivation, Test Scores, and Economic Success. *Management Science*. **58**(8) 1438-1457.

Tonin, M., Vlassopoulos, M. 2015. Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. *Management Science*. **61**(8) 1795-1811.

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

Van Dijk, F., Sonnemans, J., Van Winden, F. 2001. Incentive Systems in a Real Effort Experiment. *European Economic Review*. **45**(2) 187-214.

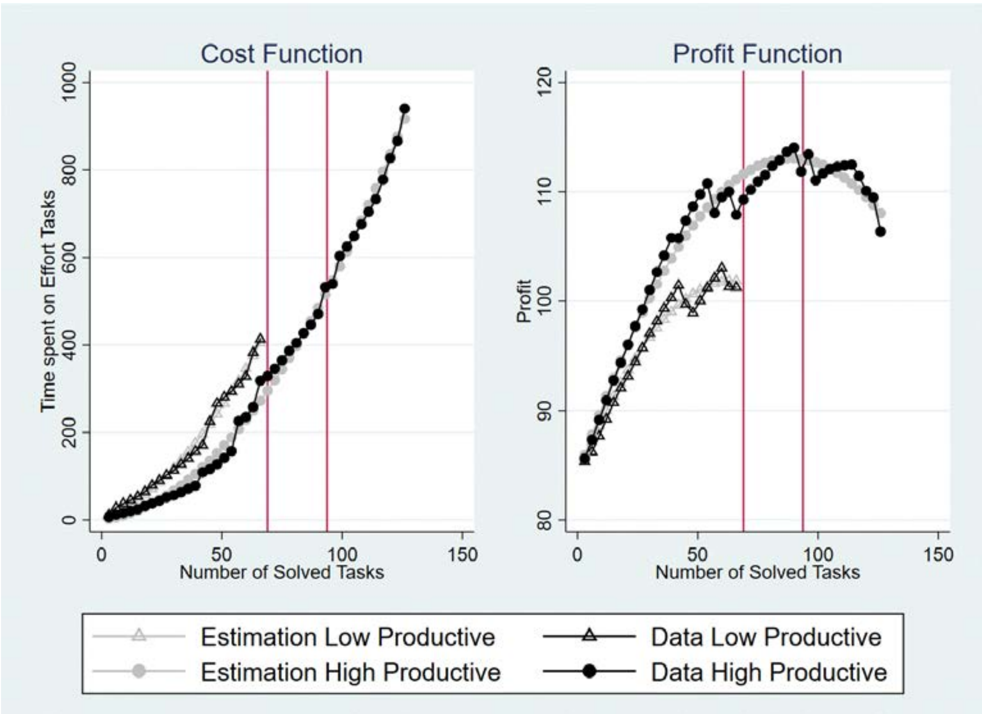
Zizzo, D.J. 2010. Experimenter Demand Effects in Economic Experiments. *Experimental Economics*. **13**(1) 75-98.

Appendix A 1: Estimating money-maximizing output

In section 3, we estimate the benchmark for money-maximizing output by assuming a quadratic cost function. We now show that this is a good approximation to observed behavior.

Figure A1 illustrates in the left panel that our estimated quadratic cost functions match observed data closely for two selected workers (one high productive worker, HP, one low productive worker, LP). The LP has a steep cost function, i.e., he or she needs more time to complete a given number of tasks than a HP, and he or she stops earlier than the HP. We estimate the cost function according to eq. (3) for each worker separately using an OLS regression. The overall fit of these estimates to observed data is excellent (adjusted R^2 equals on average 0.958 for low and 0.997 for high productivity workers *Figure A1*).

Figure A1: Comparison of estimated and actual cost and profit function



Notes: The left panel shows the estimated and actual cost function of a high and a low productive participant. The right panel shows the estimated and actual profit function of a high and a low productive participant.

The right panel illustrates how observed profit relates to estimated profit for the same two workers as in the left panel. Again, we note a close match (adjusted R^2 = 0.987). In the example, the LP has a $\hat{\beta}_1$ of 2.13 and a $\hat{\beta}_2$ of 0.06 and the HP a $\hat{\beta}_1$of 0.66 and a $\hat{\beta}_2$ of 0.05, respectively. These estimates imply a money-maximizing number of completed tasks for the LP of 69 and for the HP of 93 (see vertical lines in *Figure A1*). In this example, the LP is very close to the money-maximizing output as he or she completes 66 tasks ($\Delta Q_L = 3$). In contrast, the HP has a ΔQ_H of 33.

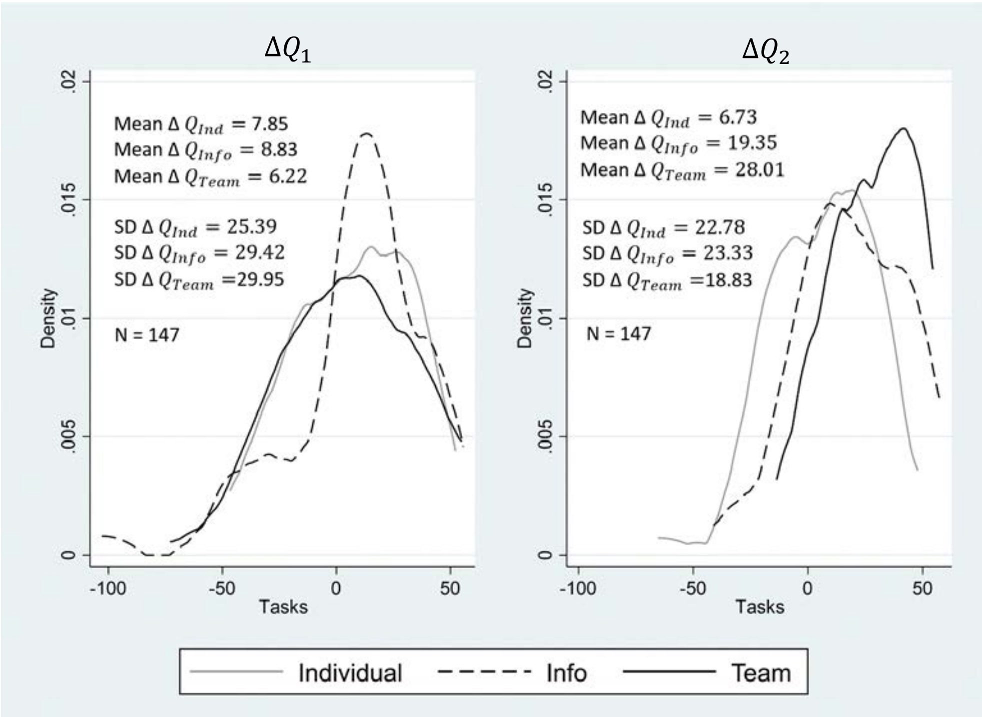
Appendix A 2: Intrinsic motivation

We develop a novel technique to measure motivation to work hard as the deviation from the money-maximizing benchmark in a real-effort experiment. Section A2.1 provides a detailed description of this measure and section A2.2 discusses potential confounds.

A2.1 Detailed description of the measure for intrinsic motivation

Figure A2 shows distributions of our motivation measures in phase 1 (ΔQ_1) and phase 2 (ΔQ_2). On average, workers are positively motivated in all treatments ($\Delta Q > 0$). We find no treatment differences for mean motivation in phase 1 (Individual vs. Info: 7.85 vs. 8.83, p = 0.818; Individual vs. Team: 7.85 vs. 6.22, p = 0.772, two-sided t-test). In the Individual treatment, ΔQ is constant over both phases as expected (ΔQ_1= 7.85, ΔQ_2= 6.73, p = 0.709, two-sided t-test) because nothing changes between these phases. The constant average excess output results because workers persistently deviate (positively or negatively) from the money-maximizing benchmark (Pearson r between ΔQ_1 and ΔQ_2 = 0.62, p = 0.000). Workers move somewhat closer to $\hat{Q}^*$ in phase 2 (note the smaller SD in the right panel), which is indicative of moderate learning effects. As predicted by assuming image concerns and prosocial preferences (see section 3), deviations from money-maximizing output increase in Info and Team from phase 1 to 2 (see dashed and solid black lines shifting right), but not in Individual.

Figure A2: Distribution of motivation measures ($\boldsymbol{\Delta Q_1}$ and $\boldsymbol{\Delta Q_2}$)



Notes: Kernel density estimates of deviations from the money-maximizing benchmark in phase 1 (ΔQ_1) and in phase 2 (ΔQ_2), in all treatments. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.

A2.2 Testing for potential confounds

Our measure of motivation is based on deviations from optimal output for a money maximizer. We interpret these deviations as being the result of a deliberate choice of workers who like (or dislike) the task. However, deviations from the optimum may not necessarily be due to maximization of a utility function, they may also be due to behavioral noise or systematic failure to maximize. We also consider potential bias in our estimation of $\hat{Q}^*$ as a source of confound. We now argue that these confounds do not cause mismeasurement.

Motivation is an important driver of output. Section A2.1 has shown that individual deviations from the money-maximizing benchmark are highly correlated across phases, meaning that deviations are systematic rather than purely random. *Table A1* uses OLS regression and dominance analysis for the Individual treatment only (because it is the only treatment with constant incentives) to check the explanatory power of intrinsic motivation with regard to (variations in) output in phase 2.[17] Column (1) shows that work motivation in phase 1 predicts output in phase 2 rather well. In particular, if ΔQ_1 increases by one standard deviation (25.39, see *Figure A2*), output in phase 2 increases by 11 tasks. As expected, an increase of optimal output (which depends on productivity) significantly increases output in phase 2.

Table A1: Motivation is an important driver of output

	(1)	(2)
	Output Phase 2	Dominance Statistic
ΔQ_1	0.45***	0.14
	(0.09)	
$\hat{Q}_2^*$	0.66***	0.44
(Money-Maximizing output in phase 2)	(0.09)	
_cons	33.94***	
	(8.40)	
Adjusted R^2	0.58	
N	50	50

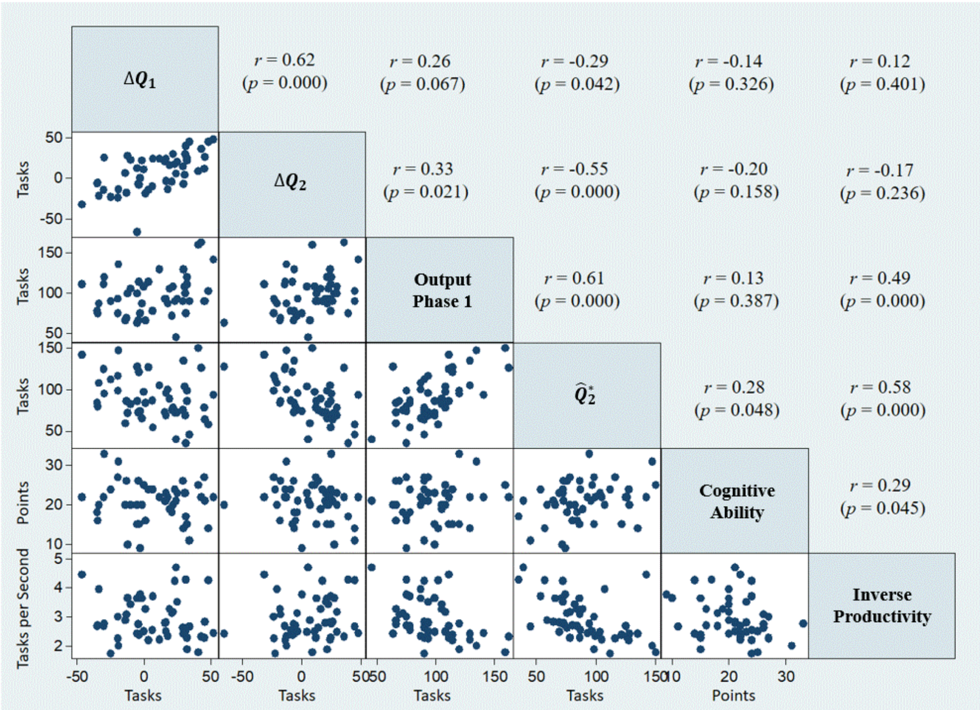
Notes: OLS Regressions and dominance analyses with the number of completed tasks in phase 2 as dependent variable. One subject was omitted as he or she completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.. Numbers in parentheses indicate standard errors. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

[17] We use a program developed by Luchman (2013). The objective of dominance analysis (DA) is to build weighted averages of the contribution to the overall fit of the model for each variable. See Gromping (2007) for a discussion of DA and Budescu (1993) for a more technical explanation.

Column (2) presents the results of a dominance analysis which shows the relative contribution of each independent variable in explaining variation of output in phase 2. In particular, 44% of the variance in output is explained by money-maximizing output in phase 2 and 14% of the variation is explained by our motivation measure (ΔQ_1). While randomness is clearly present, the fact that our motivation measure explains a considerable share of the variance of output in phase 2 indicates that intrinsic motivation is an important driver of output.

Deviations from the money-maximizing benchmark are deliberate. Deviations from money-maximizing output may be due to systematic mistakes in optimization. For example, workers may simply be unable to calculate the optimal output. We think this confound is unlikely to matter in our experiment because we explained the optimal switching point prominently in our instructions (see appendix C). While doing so is likely to have induced an experimenter demand effect, we believe that providing this information is unproblematic in our context. The reason is that we are interested in *deliberate* deviations from optimal output.

***Figure A3*:** (Pearson) Correlation matrix



Notes: N = 50.The cells with numbers in the upper right half of the table show the Pearson correlation between two variables, the scatterplots show the corresponding distribution. For example, the Pearson correlation between ΔQ_1 and ΔQ_1 is r = 0.62 (see second cell in first row) and the corresponding distribution of deviations is shown in the second cell in the first column.

In line with our belief, we find no significant relation between the score of the cognitive ability test and our measures of motivation $\Delta Q_1$and ΔQ_2, respectively (Pearson r = -0.14, p = 0.326, for ΔQ_1& r = -0.20, p = 0.158, for ΔQ_2, see second column from the right in *Figure A3*). While cognitive ability is not related to motivation, it is, in turn, positively related to the optimal output in phase 2 (p =

0.048). Taken together, these correlations indicate that more cognitively able workers are more productive but not more motivated to work beyond what is optimal for a money maximizer.

Estimate of money-maximizing output is not biased. In section 3, we assume that work motivation does not affect how fast but only how long a worker completes tasks, because non-monetary effort cost depending on the speed of work are negligible and incentives are such that a rational, negatively motivated worker should work fast, but switch to leisure earlier than predicted by the money-maximizing benchmark. The reason is as follows. The earnings of a worker depend on the payoff he or she gets through calculating tasks and on the time left for leisure. The number of tasks completed can be influenced in two ways: a worker can vary the calculation speed ("intensive margin") or the time spent in the effort task ("extensive margin"). If a worker does not like calculating cross sums and decides to work more slowly (reduce effort intensity), he or she reduces his or her payoff in two ways: first, fewer tasks are completed, which reduces money earned during the effort task as workers are paid a piece rate. Second, less money is earned in leisure as calculating slowly takes more time than is optimal and reduces the time spent in leisure.

Yet, if motivated participants were more productive (i.e., $\partial\beta_{1,2}/\partial\mu < 0$), the optimal output of a money maximizer would increase in motivation ($\partial Q^*/\partial\mu > 0$) and we would thus underestimate motivation for the positively motivated and overestimate motivation for the negatively motivated workers. While we cannot test $\partial\beta_{1,2}/\partial\mu \neq 0$ directly, we can test the effect of a change in monetary incentives w/l on productivity as a proxy. If cutting the piece rate by 43% (see *Table A2*) does not affect productivity, we are confident that motivation does not affect productivity either.

Table A2: Main phases of additional control treatments

Treatment	**Phase 0** (5 minutes)	**Phase 1** (20 minutes)	**Phase 2** (20 minutes)
Control A (*N* = 24)	Player *i* gets 0.7 points per task completed by *i*	Same as Phase 0 in Control A **&** Opportunity cost of work (1 point per 15'' of leisure)	Same as Phase 1 except that Player *i* gets 0.4 points per task completed by *i*
Control B (*N* = 48)	Player *i* gets 0.4 points per task completed by *i*	Same as Phase 0 in Control B **&** Opportunity cost of work (1 point per 15'' of leisure)	Same as Phase 1 except that Player *i* gets 0.7 points per task completed by *i*

To test, we run two additional treatments described in *Table A2*. Control A (*N* = 24) is the same as treatment Individual in *Table 1* except for phase 2 where the piece rate drops from 0.7 to 0.4 points per task. Control B (*N* = 48) is the same as control A except that the piece rate is low (0.4 points) first and increases to 0.7 in phase 2. All treatments were preceded by a test for cognitive ability (20 minutes) and an unpaid trial (4 minutes), and followed by a survey on socio-demographics. We measure productivity based on the number of tasks completed in phase 0 as well as average time per screen for the first ten screens in phases 1 and 2. *Table A3* shows that there is significant learning in both treatments as the average time per screen for the first ten screens decreases significantly between phases (Control A: $p = 0.019$; Control B: $p = 0.013$). This learning effect, however, does not differ between piece rates ($p = 0.326$). Additionally, we compare output in phase 0 to show that productivity is not affected by the piece rate ($p = 0.175$).

Table A3**:** Productivity does not change with the piece rate

		Change in Inverse Productivity (between phase 1 and 2)	Output Phase 0
Mean	Control A ($N = 24$)	-0.42 ($p = 0.019$, one sample t-test)	55.00
	Control B ($N = 48$)	-0.24 ($p = 0.013$, one sample t-test)	58.65
p – value Two-sided t-test		0.326	0.175

Notes: We use inverse productivity measured by the average time per screen for the first ten screens in phases 1 and 2 as a proxy for productivity.

Table A4: Productivity does not differ between control treatments

	Inverse productivity Phase 2
Control B	-0.06
	(0.12)
High cognitive Ability	-0.23**
	(0.11)
Inverse Productivity Phase 1	0.51***
	(0.06)
_cons	1.39***
	(0.24)
R^2	0.59
N	67

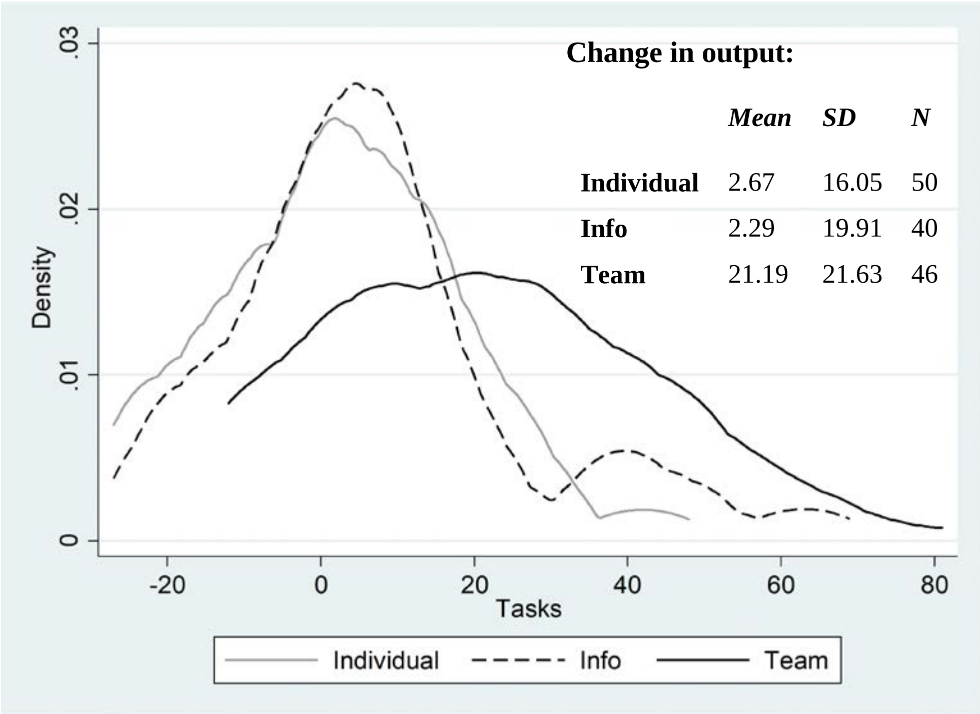
Notes: OLS-Regression analyzing what determines inverse productivity measured as average time per screen in the 1st 10 screens in phase 2. Five subjects had to be omitted as the inverse of the time it takes them to complete the first 10 screens could not be calculated in phase 1 or 2. Numbers in parentheses indicate standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A4 shows an OLS-Regression to identify the main explanatory variables influencing productivity in phase 2. We use productivity in the first 10 screens in phase 2, as there is a change in wage in this phase. If wage influenced productivity, we would see a significant effect of our dummy variable for the treatment (*Control B*). We find no effect for monetary incentives on productivity (see variable *Control B*) and conclude that it is unlikely that motivation affects productivity.

Appendix B 1: Aggregated treatment effects

This section discusses the aggregated effect of team incentives and observation. *Figure B1* shows the change in the number of completed tasks between phases 1 and 2 for all treatments separately. While we find little evidence for peer pressure induced by observation (see dashed line), the introduction of team incentives increases output significantly (see solid black line).

Figure B1: Treatment effects on output



Notes: Kernel density functions of the change in output between phases for the different treatments. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.

Table B1 shows the results from OLS regressions including a dummy variable indicating the different treatments. In phase 1, workers in all treatments receive the same incentives, but workers in Info and Team know that their output will be reported to their team members after the first phase, which could affect their performance. However, the groups do not differ significantly in any of the dependent variables discussed (see *Info* and *Team* in columns 1-3). Looking at the interaction terms between treatments and phase 2 (the phase in which the treatments were introduced), there is a significant combined effect of observation and team incentives (see variable *Phase 2 x Team*) on time spent on the task and output, but no effect of observation alone (with the exception of a weakly significant effect on worktime, see variable *Phase 2 x Info* in columns (1)-(3)). To isolate the additional effect of team incentives we use the Info treatment as the baseline in columns (4)-(6). As expected from the result described above, team incentives significantly increase worktime and output controlling for observation.

Table B1: Workers in the Team treatment increase effort, while observation alone has only small effects

	Baseline: Individual			Baseline: Info		
	(1)	(2)	(3)	(4)	(5)	(6)
	Worktime	Output	Inverse Productivity	Worktime	Output	Inverse Productivity
Phase 2	-3.96	2.76	-0.21**	91.36**	7.29**	-0.30***
	(27.88)	(2.35)	(0.09)	(39.09)	(2.75)	(0.09)
Info	27.04	1.31	0.15			
	(55.85)	(5.77)	(0.15)			
Team	-39.83	-7.71	0.19	-66.87	-9.02	0.03
	(50.69)	(5.31)	(0.18)	(55.20)	(5.38)	(0.20)
Phase 2 x Info	95.32*	4.53	-0.09			
	(47.88)	(3.61)	(0.13)			
Phase 2 x Team	243.21***	18.43***	-0.05	147.89***	13.90***	0.03
	(41.36)	(3.41)	(0.13)	(49.70)	(3.71)	(0.13)
_cons	640.28***	93.96***	2.93***	667.32***	95.27***	3.08***
	(36.52)	(4.05)	(0.09)	(42.44)	(4.13)	(0.12)
R^2	0.09	0.06	0.03	0.10	0.08	0.03
Subjects	147	147	147	97	97	97
Observations	294	294	294	194	194	194

Notes: OLS regression with standard errors clustered by team for time spent working, output and inverse productivity (measured as the time per screen for the first 10 screens). The sample includes 153 subjects. As each subject appears twice (in phase 1 and phase 2), this would lead to 306 observations. 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix B 2: Heterogeneity in the effectiveness of team incentives combined with observation by peers

Section 4 discusses moderating effects of motivation on team incentives and observation. To do so, we split the participants in terciles based on their deviation from money-maximizing output in phase 1, which are shown in *Table B2*. The terciles overlap slightly, because we have built the terciles separately for each treatment in order to have an equal share of each motivation type in each treatment. We test the teciles for differences in intrinsic motivation in phase 1 between treatments, which could confound our treatment effects and do not find any significant differences using an OLS regression with clustered robust standard errors ($p > 0.1$ for all treatment and tercile comparisons, Wald tests).

Table B2: Motivation terciles

ΔQ (in tasks)	Negative Motivation (N = 50)	Money Maximizer (N = 49)	Positive Motivation (N = 48)
Mean	-23.35	9.99	37.53
Median	-21.01	11.45	36.14
SD	20.11	7.85	10.80

Notes: Terciles over all treatments used to build equally sized motivation types. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output.

Table B3 tests the terciles for differences in intrinsic motivation in phase 1 between treatments. We do not find any significant differences using an OLS regression with clustered robust standard errors ($p > 0.1$ for all treatment and tercile comparisons, Wald tests).

Table B3: Average intrinsic motivation in each motivation tercile does not differ between treatments.

	ΔQ in Phase 1		
	Negative Motivation	Money Maximizer	Positive Motivation
Info	-1.62	2.63	2.24
	(7.67)	(2.69)	(4.05)
Team	-6.16	-3.29	2.92
	(4.70)	(3.26)	(4.22)
_cons	-20.83***	10.21***	35.81***
	(2.57)	(2.52)	(2.33)
R^2	0.02	0.10	0.01
N	50	49	48

Notes: Splitting the subjects into terciles results in bins of 51 observations each. Six subjects are omitted as they completed less than 30 tasks, which is insufficient to precisely predict money-maximizing output. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B4: Motivation moderates team incentives combined with observation

	Change in Output	
	(1) Baseline: Individual	(2) Baseline: Info
Info	3.19	
	(5.26)	
Team	4.13	0.94
	(4.29)	(4.19)
Negative Motivation	12.50**	20.07***
	(5.73)	(6.21)
Money Maximizer	2.79	-1.31
	(4.85)	(5.83)
Info x Negative Motivation	7.58	
	(8.43)	
Info x Money Maximizer	-4.10	
	(7.57)	
Team x Negative Motivation	19.00**	11.43
	(8.04)	(8.40)
Team x Money Maximizer	24.21***	28.31***
	(6.05)	(6.87)
_cons	-2.44	0.75
	(3.78)	(3.67)
R^2	0.39	0.41
N	147	97

Notes: OLS Regression with robust standard errors clustered by teams using a diff-in-diff approach to identify interaction effects between intrinsic motivation and observation 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 standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B4 shows an OLS regressions including interaction terms for motivation types and treatment. We expect to find that the treatment effect decreases in motivation (see eq. 9). Starting with the combined effect of observation and team incentives in column (1), we find support for this hypothesis (see *Team x Negative Motivation*: + 19.00 tasks, $p = 0.022$, controlling for learning effects). We find no significant interaction between observation and intrinsic motivation see *Info x Negative Motivation*: + 7.58 tasks, $p = 0.373$, controlling for learning effects). To analyze the additional effect of team incentives controlling for information, we use the Info treatment as a baseline in column (2). We find that while observation only affects negatively motivated workers, team incentives also affect money maximizers (see *Team x Negative Money Maximizers:* +28.31 tasks, $p = 0.000$).

Table B5: Teams work better for the highly productive ($\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$)

	Change in Output	
	(1) Baseline: Individual	(2) Baseline: Info
Info	22.51**	
	(10.84)	
Team	43.03***	20.52
	(11.52)	(12.35)
$\hat{\beta}_1$	3.88	3.90
	(2.44)	(2.43)
$\hat{\beta}_2$	143.04	-149.91
	(91.00)	(123.28)
Info x $\hat{\beta}_1$	0.02	
	(3.44)	
Team x $\hat{\beta}_1$	-3.89	-3.91
	(3.89)	(3.89)
Info x $\hat{\beta}_2$	-292.95*	
	(152.87)	
Team x $\hat{\beta}_2$	-319.43**	-26.48
	(135.75)	(159.43)
_cons	-9.78	12.73
	(7.02)	(8.29)
R^2	0.21	0.18
N	147	97

Notes: OLS Regression with robust standard errors clustered by teams using a diff-in-diff approach to identify interaction effects between productivity estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ and team incentives as well as observation. Six subjects are omitted as they 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 B5 shows an OLS regression with standard errors clustered by team and interaction terms for productivity (measure as $\hat{\beta}_1$ and $\hat{\beta}_2$) and treatments to analyze if productivity shapes the effectiveness of team incentives and observation.[18] In column (1) we use the Individual treatment as a baseline and find evidence for a moderating effect of productivity on team incentives but not on observation (see *Team x* $\hat{\beta}_2$ and *Info x* $\hat{\beta}_2$*)*. An increase in $\hat{\beta}_2$ by one standard deviation decreases the effect of team incentives combined with observation by 8.91 tasks (difference-in-difference between Individual and

[18] Note that higher parameters mean that an individual is less productive as it takes him or her more time to complete a screen.

Team, p = 0.023). In column (2), we isolate the effect of team incentives by using Info as the baseline treatment and find no significant differences (see *Team x* $\hat{\beta}_2$, p = 0.869), indicating that the productivity shapes the effect of team incentives combined with observation by moderating the effect of observation.

Table B6: Teams work better for the highly productive (Time per screen)

	Change in Output	
	(1) Baseline: Individual	(2) Baseline: Info
Info	4.83	
	(8.45)	
Team	38.76***	33.94***
	(10.43)	(10.48)
Inverse Productivity	2.69	2.45
	(2.21)	(1.93)
Info x Inverse Productivity	-0.23	
	(2.93)	
Team x Inverse Productivity	-6.69*	-6.46*
	(3.42)	(3.26)
_cons	-5.10	-0.28
	(5.96)	(6.01)
R^2	0.16	0.13
N	147	97

Notes: OLS Regression with robust standard errors clustered by teams using a diff-in-diff approach to identify interaction effects between inverse productivity and team incentives as well as observation. Inverse productivity is proxied by the average time per screen for the first ten screens in phase 1. Six subjects are omitted as they 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 B6 replicates the results of *Table B5* using another proxy for productivity. The reason is that, $\hat{\beta}_1$ and $\hat{\beta}_2$ are, by construction, correlated with our measure of intrinsic motivation ΔQ_1. Therefore, we use the time it takes participants to complete the first 10 screens in phase 1 as a proxy when we control for productivity. This proxy is positively correlated with $\hat{\beta}_1$ and $\hat{\beta}_2$ ($\hat{\beta}_1$: Pearson r = 0.58, p = 0.000; $\hat{\beta}_2$: Pearson r = 0.25, p = 0.002), but not with ΔQ_1 (Pearson r = 0.01, p = 0.912). Furthermore, output in phase 1 is strongly correlated to this productivity measure (Pearson r = -0.44, p < 0.000) and the time it takes participants to complete the first 10 screens in phase 1 does not differ between treatments (Individual vs. Info: p = 0.727 and Individual vs. Team p = 0.804, Fisher's z-test). Using this measure of productivity, we find again that more productive workers react more strongly to team incentives combined with observation (*see Team x Inverse Productivity*).

Appendix B 3: Ranking

To test if reciprocity concerns can explain our result that motivation shapes the effectiveness of team incentives and observation, in section 4.3 we check if ranks shape the effect of the treatments as well. On the one hand relative performance, i.e. ranks, depends on the productivity of the participant. On the other hand, participants can increase the number of completed tasks by spending more time on the real effort tasks, which is represented in our motivation measure. *Table B7* shows that about 59 % of the participants who are positively motivated achieve rank 1, compared to 16% achieving rank 3. This picture is reversed looking at negatively motivated participants, as for them the share of rank 1 participants is 21 % whereas the share of rank 3 participants is to 36%.

***Table B7*:** Motivation determines relative performance

	Rank 1	Rank 2	Rank 3	Total
Negative Motivation	7 (21.21%)	14 (42.42%)	12 (36.36%)	33 (100%)
Money Maximizer	7 (21.88%)	12 (37.50%)	13 (40.63%)	32 (100%)
Positive Motivation	19 (59.38%)	8 (25.00%)	5 (15.63%)	32 (100%)
Total	33	34	31	97

Notes: Frequency table for motivation terciles and ranks in phase 1 including the Info and the Team treatment. Five subjects are excluded as they completed less than 30 tasks, which is insufficient to predict money-maximizing output or productivity.

To account for the importance of productivity types, *Table B8* shows an ordered logit regression with rank as dependent variable for the Info and the Team treatment. We find that the chance of achieving a larger rank (which is equivalent to a weaker relative performance, as rank 3 represents workers with the lowest number of completed tasks in their team) increases with the time per screen for the first ten screens and therefore is higher for less able workers (see variable *Inverse Productivity*) as well as for negatively compared to positively motivated workers (see variable *Negative Motivation*). We use inverse productivity rather than our estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$ to avoid mulitcollinearity (see discussion in appendix B2).

Table B9 shows different ranks separately. As there are no teams in the individual treatment, we randomly match workers into teams in the Individual treatment. We repeat this procedure 100 times. The results presented below are averages over these 100 matchings. In column 1, we find that participants in the Team and the Info treatment significantly increase output in phase 2 compared to the Individual treatment. However, for workers achieving rank 2 or 3, we do not find an effect of the Info treatment, but the workers in the Team treatment increase output in phase 2 more strongly than in the Individual treatment for all ranks.

Table B8: Rank increases in inverse productivity and decreases in motivation

	Rank
Inverse Productivity	2.15**
	(0.65)
Money Maximizer	4.37**
	(2.87)
Negative Motivation	5.51***
	(3.22)
Cut1 _cons	2.64***
	(0.96)
Cut2 _cons	4.45***
	(1.08)
N	97

Notes: Ordered logit regression with rank as dependent variable for Info and Team treatment. Coefficients are represented as odd ratios. As there is no ranking in the Individual treatment, we only consider subjects in the Info and Team treatment. Five subjects are excluded as they completed less than 30 tasks, which is insufficient to predict money-maximizing output or productivity. Numbers in parentheses indicate robust standard errors. Stars indicate significance at the following levels: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table B9: All ranks increase output in Team

	Change in Output (average of 100 random matchings in the Individual group)		
	(1) Rank 1	(2) Rank 2	(3) Rank 3
Info	9.30*	3.80	0.44
	(4.66)	(7.06)	(6.62)
Team	10.34**	25.15***	19.44***
	(4.56)	(6.72)	(6.37)
_cons	-4.72	3.08	10.36
	(2.98)	(3.94)	(3.91)
R^2	0.11	0.23	0.20
N	50	51	46

Notes: Change in output between phases based on treatment for different ranks. The Individual treatment is used as the baseline group. Splitting the 153 subjects in three groups, leads to a group size of 51 for each rank. Six subjects are omitted as they 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 B10: Relative performance does not moderate team effects combined with observation

	Change in Output	
	(1) Baseline Individual	(2) Baseline Info
Info	9.30*	
	(4.70)	
Team	10.34**	1.04
	(4.59)	(5.01)
Rank2	7.80	2.29
	(4.87)	(7.56)
Rank3	15.08**	6.21
(weakest relative performance)	(4.87)	(6.01)
Info x Rank 2	-5.50	
	(8.98)	
Info x Rank 3	-8.86	
	(7.74)	
Team x Rank 2	14.81	20.32*
	(8.16)	(10.00)
Team x Rank 3	9.10	17.96**
	(8.01)	(8.75)
_cons	-4.72	4.59
	(2.30)	(3.62)
R^2	0.23	0.24
N	147	97

Notes: Mean effect of 100 OLS Regressions (random team matching in Individual treatment) using a diff-in-diff approach to identify interaction effects between ranking and the change in output. Column (1) uses Individual treatment as baseline group, whereas the Info treatment is used as a baseline in column (2). Six subjects are omitted as they 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

To control for learning effects, we include treatments and ranks in *Table B10*. In column 1, we use the Individual treatment as a baseline and do not find evidence that ranks shape the effectiveness of team in incentives and observation or observation alone Team treatment (*Team x Rank 2* & *Team x Rank 3* and *Info x Rank 2* & *Info x Rank 3,p* > 0.1). In column (2), we use the Info treatment as the baseline and find that participants with a weaker relative performance increase output more if team incentives are in place. This indicates that there is a runoff for the best rank in the Info treatment, which is counterbalanced by increasing motivation for rank 2 and 3 workers in the Team treatment (see variables *Team x Rank 2* & *Team x Rank 3* in column 2).

Appendix B 4: Do motivated participants have weaker social preferences?

As shown in section 5, we find that team incentives and observation by peers lead to an increase in output of 25 percent for the average worker. We show that this increase is mainly driven by workers with negative or no intrinsic motivation. We suggest that this is so because of cost considerations. In particular, we argue that expanding output is profitable for those with output below the money-maximizing output (i.e., those with negative intrinsic motivation) but expensive for those beyond the optimum (i.e., those with positive intrinsic motivation). A concern with this argument is that the responsiveness to team incentives depends on workers cooperativeness. If those with positive intrinsic motivation were less cooperative than those with negative intrinsic motivation, this would also explain our results.

To test for this concern, we ran two control sessions in which participants play a public good game either before or after the real effort tasks (see experimental design in section 2).[19] As suggested by Fischbacher, Gächter and Fehr (2001), participants make two contribution decisions, first they make an unconditional decision and second they make a conditional decision. The first choice is supposed to measure cooperativeness, the second conditional cooperation. We find no relation between these two measures and intrinsic motivation.

To be more precise, we find that unconditional contribution to the public good and intrinsic motivation in phase 1 have a Pearson correlation coefficient of 0.012 ($p = 0.935$, $N = 52$). To investigate the relation between conditional cooperativeness and motivation, we follow Fischbacher et al. (2001) and classify workers into types: conditional cooperators, selfish types and others. We test whether the share of cooperation types differs across motivation terciles shown in *Table B2*. According to a Fisher test, shares do not differ across types ($p = 0.391$). We conclude that differences in cooperativeness between motivation types cannot explain the moderating effect of motivation.

[19] In total 56 subjects participated in the control session. However, three subjects are excluded as they signaled a computer problem to the experimenter. One more subject is excluded as he or she did not complete any tasks in phase 1 and we can therefore not estimate money-maximizing output.

Appendix C: Instructions for participants

(Instructions were handed out separately at the beginning of every phase)

(Instructions are for the treatment Team)

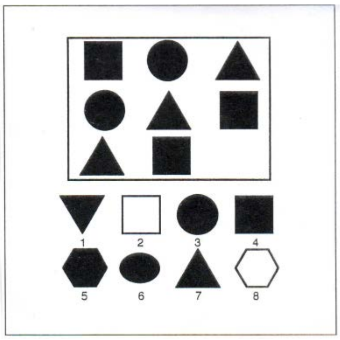
General explanation: You are now taking part in an economics experiment. During the experiment you can earn money. Therefore it is important that you read the following instructions carefully.

These instructions are solely for your private information. Do not communicate with other participants during the experiment! If you have any questions, please raise your hand and an assistant will help you.

During the experiment your earnings will be calculated in points. Your earnings from the experiment will be paid to you in cash right after the experiment. Your earnings in points will be converted into cash according to the following exchange rate: **15 Points = 1 EUR.** The experiment consists of several phases.

Explanation for the 1st phase: You will see 36 screens in sequence. On each screen you will find a task. We give you 20 minutes and we ask you to solve as many tasks as possible correctly.

Example of a task: Here is an example of a task. The correct solution of this task is symbol #3. There is a square, a circle, and a triangle in each row. In the third row, there is a triangle and a square, but no circle. Therefore, symbol #3 is the correct answer.



Explanation for the 2nd phase

What to do in the 2nd phase: In this phase you will calculate cross-sums. You have to sum up the sequence of digits. Here is an example: 5 7 8 0 3

Your task is to calculate the cross-sum, that is: 5+7+8+0+3. The correct answer in this example is 23.

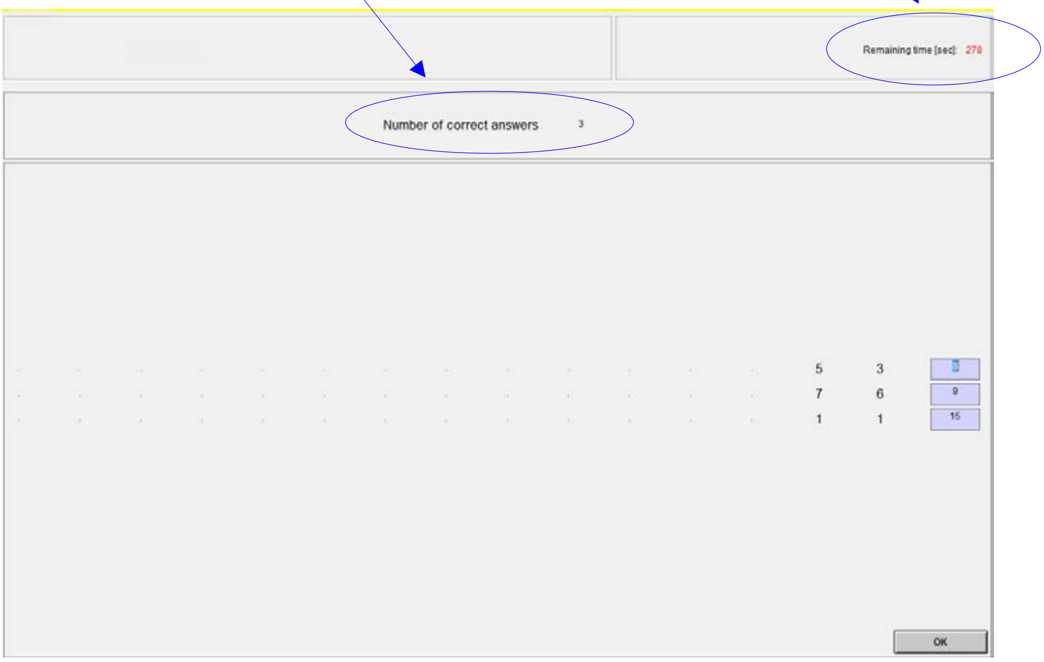
How your earnings are calculated: The 2nd phase lasts for 5 minutes (300 seconds). For each correct cross-sum you receive a payment of **0.7 points.**

How the second phase works: You solve cross-sums on screens (you see an example in the next pages). Each screen contains 3 cross-sums. Only if you solve all three calculations correctly and click the "OK"-button, the next screen will be shown. If you make a mistake in one of the calculations, the program tells you where you made the mistake and you have to revise your answer (look at the bottom left corner).

In the beginning the cross-sums consist of 2 digits. After every 5 screens one digit is added to each cross-sum on the screen; i.e., the 3 cross-sums from the 6th screen each consist of 3 digits, from the 11th screen of 4 digits, from the 16th screen of 5 digits, and so on.

Warning: when you enter a new screen of cross-sums the answers to the cross-sums from the previous screen of cross-sums will still be in the **answering fields** (like on the screen on the next page); these have to be **overwritten**. You can use the mouse or the TAB-button to manoeuvre from one cross-sum to the next. However, for many people the easiest way to manoeuvre from one cross-sum to the next cross-sum on the screen is by using the TAB-button,→ and only using the mouse to click the "OK"-button.

What the screens look like: On the top of the screen you see the **remaining time** to make your calculations and the **number of correct answers**. Note that as you press “OK” and start a new set of cross sums, the answers to the previous cross sums remain.



At the end of phase 2 you will be informed about how many cross-sums you have solved and your earnings in this particular phase.

Training phase: Before entering phase 2, we conduct a 4 minute training phase to familiarize you with calculating cross-sums and entering your decisions on the computer. Compared with phase 2 there are the following two differences:

1. The difficulty of the cross-sums increases after every two screens of cross-sums (6 cross-sums) instead of after every five screens. This way you can experience the difference in difficulty of cross-sums with different numbers of digits.
2. You will not earn any money in the training phase. However, previous studies have shown that people become considerably faster at doing cross-sums when being familiar with them (and thus increase their potential earnings of the future phases of the experiment) and we therefore strongly encourage you to do so.

Explanation for the 3rd phase

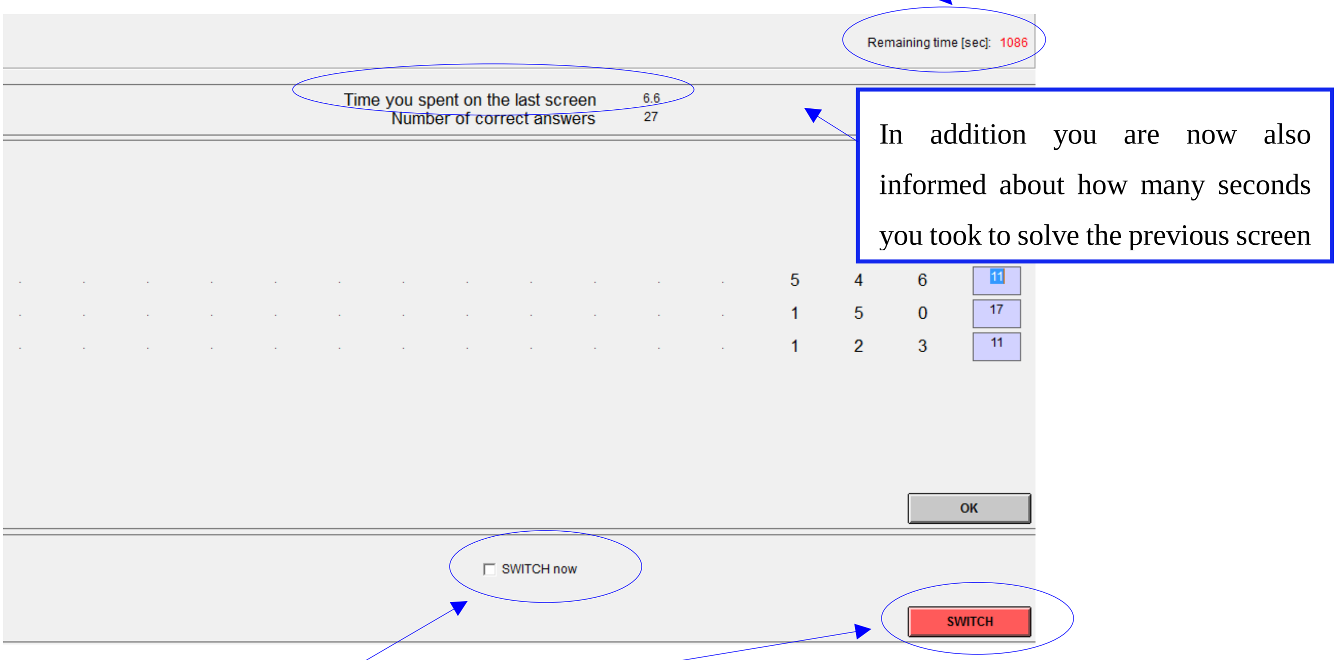
What to do in the 3rd phase: In the 3rd phase, you have to solve the same kind of tasks as in phase 2, but an additional task, the SWITCH-Task is introduced and your payment is calculated differently. In the SWITCH-Task you don’t have to calculate cross-sums anymore.

How your earnings are calculated: The 3rd phase lasts for 20 minutes (1200 seconds). We will now explain how your earnings are determined.

1. Your earnings from the solution of tasks: for each correct cross-sum you receive **0.7 points.**
2. Your earnings from the SWITCH-Task: during phase 3 you can switch to the so called SWITCH-Task at any time. In the remaining time after switching to the SWITCH-Task until the end of the phase you earn **1 point for each 15 seconds**.

Example for earning from the SWITCH-Task: This phase lasts for 1200 seconds. Suppose you switch to the SWITCH-Task after 800 seconds. This means that 400 seconds remain till the end of the round. For these 400 seconds you earn **26.6 (400/15) points.**

What the screens look like: On the top of the screen you see the **remaining time** to make your calculations and the **number of points earned** so far. Note that as you press "OK" and start a new set of cross-sums, the answers to the previous cross-sums remain.



SWITCH-Task: If you press the **"SWITCH"-button** on any of the screens, you switch to the SWITCH task. Remember that you earn **1 point** for each **15 seconds** you have spent on the SWITCH task.

To ensure that you do not switch to the SWITCH task by accident you have to mark the **"SWITCH" checkbox** before you can press the "SWITCH"-button.

Important: Once you have switched to the SWITCH task you **cannot switch back** to doing cross-sums for the remainder of the phase.

Once you have switched to the SWITCH task, you will wait until the phase ends; that is, the 20 minutes are up. If you are in the SWITCH task, you must remain seated and you are not allowed to communicate with others.

At the end of phase 3 you will be informed about how many cross-sums you have solved and your earnings in this phase.

Important: As you already know, each screen consists of 3 cross-sums, each worth 0.7 points. Therefore you can earn **2.1 points** for each screen. In the SWITCH-Task, you earn **1 point for each 15 seconds.**

This means: if you are fast in solving cross-sums, you earn more in the cross-sum task than in the SWITCH-Task. On the contrary, if you are slow in solving the cross-sums, you earn less than in the SWITCH-Task. Concretely, you receive higher earnings in the SWITCH-Task, if it takes you more than **2.1 × 15 = 32 seconds (**rounded**)** to solve the 3 cross-sums on one screen.

Number of cross-sums within your group: In this phase, you will be divided in groups of 3; this means that you will be together with 2 other persons in a group. Each of you gets a GroupID (from 1 to 3) in order to be identifiable within the group. You will be informed about your GroupID soon. You will stay in the same group and all group members will keep the same GroupIDs until the end of the experiment.

At the end of Phase 3, we will inform you about the number of cross-sums solved by the members of your group. In this way, you know who has solved the highest, the middle, and the lowest number of cross-sums in your group.

The following screen shows an example of the information you will see. Assume that you are GroupID 1 and solved 96 cross-sums. In this case, you know that you rank second and that GroupID 2 has achieved 63 and GroupID 3 has achieved 102 cross-sums.

Explanation for the 4th phase

<u>What to do in the 4th phase:</u> The structure of phase 4 is the same as in phase 3. You also stay in the same groups as before. Only your earnings are calculated differently.

<u>How your earnings are calculated:</u> The 4th phase lasts for 20 minutes (1200 seconds). We will now explain how your earnings are determined.

<u>**Your earnings from the solution of tasks:**</u> for each correct cross-sum you receive **0.7 points. Additionally you receive 0.2 points for each cross sums solved by one of your group members.** Also the other two members of your group get 0.2 points each, for each cross sum you solve.

<u>Example for earnings from the solution of tasks:</u> Assume you solve 100 cross sums, one of you group members solves 120 and the other one 80 cross sums. Then your earnings from the solution task are:

0.7 x 100 + **0.2 x 120** + **0.2 x 80** = **110**

Your points + points you get from group member2 + points you get from group member3

On the other hand, group member2 would earn 0.7 x 120 + 0.2 x 100 + 0.2 x 80 = 120, and group member 3 would earn 0.7 x 80 + 0.2 x 100 + 0.2 x 120 = 100, respectively.

The following table sums the results up:

Nr. of cross sums you solved	Nr. of cross sums solved by member 2	Nr. of cross sums solved by member 3	Your earnings	Earnings member 2	Earnings member 3
100	120	80	110	120	100

<u>**Your earnings from the SWITCH-Task:**</u> as in phase 3, you can switch to the so called SWITCH-Task at any time. In the remaining time after switching to the SWITCH-Task until the end of the phase you earn **1 point for each 15 seconds**. (Your group members do not get money while you are in the SWITCH-Task)

<u>What is different to phase 3:</u> Please note that with each solved cross sum your group members earn 0.2 extra points and so do you for each of their cross sums, meaning that with each screen you solve the group earns 3 x 0.7 + 2 x (3 x 0.2) = **3.3 points**. Therefore, your group as a whole can increase total earnings if each member works up to **3.3 x 15 = 50 seconds (**rounded**)** on one screen.

This means if all group members work more than in phase 3 (up to 50 seconds per screen), the earnings off all group members can increase compared to phase 3.