Incentivizing Creativity: a Large-Scale Experiment with Tournaments and Gifts *

Christiane Bradler†, Susanne Neckermann‡, Arne Jonas Warnke§

June 2017

Abstract

This paper reports the results from a large-scale laboratory experiment that compares the impact of tournament incentives and wage gifts on a creative and a simple task. We find that tournaments substantially increase output in both tasks. By comparison, wage gifts are significantly more effective in the simple task. Additional treatments suggest that it may be the uncertain mapping between effort and output in the creative task that reduces reciprocity. Our findings provide a rationale for the frequent use of tournaments when seeking to motivate creative output.

JEL Classification: C91, D03, J33, M52
Keywords: creativity, incentives, tournament, reciprocity, experiment, crowding-out

*We gratefully acknowledge financial support from the Germany's Federal Ministry of Education and Research framework program "Economics of Science." Further, we thank the SEEK research program for their financial support as well as the many seminar and conference participants whose valuable contributions made this paper much better. Specifical thanks goes to Dirk Engelmann, Michael Kosfeld, Francois Laisney, Steve Levitt, John Morgan, Nick Zubanov, and Thomas Zwick. We thank Christian Bommer, Michael Dörsam, Radost Holler, Sascha Lehman, Johannes Moser, Vera Schmitz, Jan Schneiderwind, Mattie Thoma, Jeannine van Reeken, and Timo Vogelsang for outstanding research assistance. This paper was previously circulated as "Creativity is different: Comparing rewards across a creative and a routine task" and "Rewards and Performance: A Comparison Across a Creative and a Routine Task".

†Volkswagen, Germany. E-mail: christianebradler@yahoo.de.

‡University of Chicago, ZEW, and CREMA. E-mail: sneckermann@uchicago.edu.

§ZEW Centre for European Economic Research Mannheim. E-mail: warnke@zew.de.

1 Introduction

There is a long tradition in economics that investigates the impact of explicit incentives (e.g., tournament rewards) and implicit rewards (i.e. wage gifts) on motivation and productivity in simple and routine tasks. Overall, these studies suggest that both types of incentives have a positive effect on performance. To date, we have only very limited understanding of how performance in creative tasks responds to incentives. As the share of workers performing tasks that require them to engage in non-routine problem solving and creative thinking has increased substantially over the last several decades (Autor et al., 2003; Florida, 2002), knowing how incentives affect creative tasks has become increasingly important. One particular challenge relates to incentivizing workers to perform well in these types of jobs. The purpose of this paper is to gain a deeper understanding of 1) how creative performance responds to rewards and of 2) whether the lessons learned from simple and routine tasks can be generalized to complex tasks that have a creative component.

Towards this end, we conducted a laboratory experiment with more than 1000 subjects. In the experiment, subjects worked in groups of four agents for one principal who benefitted from the agents' efforts. Subjects worked on either a simple, routine task or on a multi-dimensional creative task that rewards the number as well as the originality of ideas, and were exposed to either a tournament incentive or a wage gift. In addition to those rewards, agents received a fixed wage.

A particular challenge is to capture creativity in a lab setting. We employ the Unusual Uses as a multi-dimensional, creative task.[1] In the task, subjects have to come up with as many and as original alternative uses for common objects such as a tin can or a sheet of paper. Hence, rather than measuring blue-sky creativity – even though Unusual Uses does allow measuring the originality of ideas as well – the task focuses on whether subjects can place common objects into a different context. This is a central element in business innovation and in corporate idea-suggestion systems(Woodman et al., 1993). In that sense, the paper complements the existing literature on creativity in economics that mostly focuses on either blue-sky creativity or on tasks that involve very little creativity such as those that involve pattern recognition. Another advantage of the task is that it measures and captures creativity along several different dimensions: quantity (the number of answers), breadth (the spread of answers across different idea categories), as well as originality (measured as either the statistical infrequency of answers or by subjective evaluation). The availability of these separate measures allows us to address issues such as quantity - quality tradeoffs when assessing the effect of incentives on creative

[1]The task is not new to the literature. In economics, it has been used by Dutcher (2012) in his study on the effects of telecommuniting on productivity. In creativity research, the task has been used by Shalley and Oldham (1997).

performance.[2]

We find a strong positive performance response to the introduction of the tournament incentive in both tasks (routine and creative). The effect sizes are of similar magnitude in the two tasks, suggesting that performance in both tasks is equally sensitive to competitive incentives. Interestingly, however, the performance response to the wage gift differs between the two tasks. Subjects in the routine task respond to the gift with an economically and statistically significant increase in their performance. The effect size is similar to that typically found in the literature on gift exchange (see, for instance, Fehr and Gächter, 2002). However, there is no statistically significant effect of the gift on performance in the creative task.[3]

One possible explanation for the absence or reduction of reciprocity in the creative task could be the employees' uncertainty about how their efforts affected the principal's payoff in the creative task (Hennig-Schmitt et al., 2010). Whereas agents had perfect control and knowledge over how many silders they positioned correctly, and, hence, how much profit they generated for the principal in the simple task, there was some uncertainty in the creative task due to the originality rating not being known to agents while they worked on the task.

To test for this explanation, we decoupled subjects' efforts from the amount of profit that they generated for the principal. In particular, we informed subjects at the end of each round about the exact number of points that their suggestions had generated during the preceding round and then allowed them to decide on how many of those points they wanted to transfer to their principal. In this set-up we find a clear reciprocal response to the wage gift in the amount of points that subjects transfer to the principal. The effect size is very similar to the behavioral response in the simple task. This suggests that there is reciprocity in creative tasks similar to what is found in simple, routine tasks when subjects have perfect control over how their actions affect the profit of the principal.

This study contributes to the small but growing literature in economics that studies the impact of rewards on creativity.[4] Previous studies have explored how creativity is influenced by the size of the reward (Ariely et al., 2009) or the type of creative task (Charness and Grieco, 2012). Laske and Schroeder (2015) focus on the multi-tasking aspect of creativity by looking at how incentives affect quantity, quality, and novelty of creative output. Erat and Gneezy (2015), by comparison, compare the effectiveness of piece rate incentives and competitive incentives and

[2]Just like any other task, this task is, however, not without limitations. Please refer to Section **??** and Section **??** for a discussion.

[3]While the coefficient is close to zero, it's the standard error is relatively large so we cannot exclude that the gift does have a small positive effect on performance. The effect size is, however, statistically significantly smaller than in the slider task.

[4]We use the term *reward* for both the tournament reward scheme as well as the wage gift as a shorthand, even though the wage gift is not a reward as it is commonly understood, i.e. rewarding past performance. Instead it is independent of both past and future performance. Fehr and Falk (2002), for example, have therefore described wage gifts as implicit rewards.

find evidence for choking under pressure.[5]

We extend this nascent economic literature on creativity as well as the literature on incentive provision in four distinct ways. First, to our knowledge this is the first study to examine the effectiveness of financial gifts for increasing creative performance. The lack of literature on this subject is surprising given both the attention that gift exchange has received in the literature in the context of incomplete contracts (see Fehr and Gächter, 2000 for an overview) and the fact that creative jobs seem to be a prime example of jobs that are complex, hard to monitor, and typically governed by incomplete contracts. Second, as far as we know, this is also the first study to compare the effect sizes of a performance-dependent, competitive incentive (tournament) with that of a performance-independent wage gift in one set-up. By doing so, our design allows a direct comparison of the cost-effectiveness of the two reward schemes. Such a comparison is especially relevant for creativity as theory suggests that these two types of rewards might affect it in fundamentally different ways (e.g., Byron and Khazanchi, 2012). Third, we also advance the existing literature by studying the response to rewards in a simple and a creative task under the same experimental conditions. This allows us to speak directly to whether or not lessons learned from simple tasks can be generalized to creative or complex tasks without having to resort to comparisons across different studies (and, hence, different experimental conditions). Fourth, we provide another data point to the discussion on whether or not it is possible to foster creative performance through financial incentives.

The paper is structured as follows. Section 2 presents an overview of the existing literature. Section 3 describes the experimental set-up, the tasks, and the treatments. Section 4 presents our main results, Section **??** investigates mechanisms and looks into a number of supplementary issues, for instance, the absence of reciprocity in the creative task, the mechanism via which tournaments increase effort, and post-treatment effects. Section 5 concludes.

2 Literature

Simple, routine tasks can be defined as tasks that "can be accomplished by machines following explicit programmed rules." (Autor et al., 2003, p. 1283). By comparison, Amabile (1997) defines creativity as the production of ideas, solutions, and products that are novel (i.e., original) and appropriate (i.e., useful) in a given situation. This presumes that there is an alternative set of tasks that allows subjects to engage in creative thinking.

[5]Another study in this realm is Eckartz et al. (2012). They implemented a creative task as well as two control tasks for comparison (Raven's IQ and a number-adding task) in one experimental set-up. They find that neither the tournament incentive nor a piece rate had any effect on performance in any of the three tasks. This makes it hard to draw clear conclusions about whether or not rewards fail to enhance creativity, since their rewards did not affect their control groups either. A likely explanation is that baseline motivation was very high in all three tasks.

The economics literatures on incentives has almost exclusively focused on routine tasks confirming what standard economic theory predicts: financial incentives have a positive effect on performance in simple tasks because agents increase effort as long as the benefits they derive from each additional unit of output exceed their effort costs. Positive incentive effects have been demonstrated for different types of performance-dependent rewards such as piece rates, where workers are rewarded according to their absolute output (for instance, Lazear, 2000) or tournaments, where workers are rewarded on the basis of their relative performance (Harbring and Irlenbusch, 2003).

Similarly, the literature on gift exchange has also almost exclusively utilized simple, routine tasks. This literature documents that explicit financial incentives are not the only way to trigger workers' performance by showing that workers reciprocate wage gifts with higher effort (Akerlof, 1982). The gift-exchange hypothesis has been tested and confirmed in a myriad of laboratory experiments with both chosen and real effort (see, e.g., Fehr et al., 1997 for an early study on the topic, or Fehr and Gächter, 2000 for an overview). There is, however, mixed evidence on the effectiveness of gift-exchange in the field (see, for example, Gneezy and List, 2006).

Taken together, these findings inform human resource management on how to optimally reward employees in jobs that involve a clearly defined and repetitive workflow. Yet it is critical to understand whether these insights into the effectiveness of different rewards also hold for jobs that have a creative component.

Performance responses to incentives may differ, for example, because there is a difference in the degree of intrinsic motivation associated with working on these tasks,[6] because creative tasks tend to be more cognitively demanding, more risky, and of less certain value than routine performances (e.g., Amabile, 1996; Eysenck, 1995). Many scholars as well as practitioners in this area have therefore suggested that motivating creative performance is fundamentally different from motivating routine performance (e.g., Amabile, 1996). The nascent literature on economics on how rewards affect creative performance was mentioned above. Yet, to date, a direct comparison of responses to rewards across the two types of tasks is missing. This paper addresses this gap.

[6] In psychology, the literature on how rewards affect creativity has centered around the question of how rewards affect intrinsic motivation – motivation to engage in an activity out of interest, enjoyment, or a personal sense of challenge (Amabile, 2013). To date, the evidence in psychology is mixed with one camp of the literature documenting effects of rewards on intrinsic motivation and creativity (e.g., Amabile, 1996; Joussemet and Koestner, 1999; Deci et al., 1999) and with a second camp documenting positive effects (e.g., Eisenberger and Rhoades, 2001; Eisenberger and Shanock, 2003). Shalley et al. (2004) and Byron and Khazanchi (2012) provide overviews of the literatures in psychology, education, and organizational studies.

3 The Experiment

In the following, we introduce the experimental tasks, the set up, the treatments, and the experimental procedures.

3.1 The Tasks

In order to assess the effectiveness of rewards for routine and creative tasks, we implemented both types of tasks in the experiment.

We use the "slider task" as a proxy for simple, routine tasks in the workplace (Gill and Prowse, 2012). The slider task is a real effort task that has a number of desirable attributes. It is easy to explain and to understand, and it does not require prior knowledge. It is identical across repetitions, involves little randomness, and leaves no scope for guessing. The task features a computer screen displaying 48 sliders on scales that range from 0 to 100. Figure 2 shows an example of the screen as it was presented in the experiment.

Initially, all sliders are positioned at zero. The aim of the task is to position as many sliders as possible at exactly 50 within 3 minutes by using the mouse. Keyboards were disconnected during the task to prevent the usage of the arrow keys. Each slider can be adjusted and re-adjusted an unlimited number of times. While moving the mouse, subjects cannot be sure whether they positioned the slider at exactly 50. The exact position of the slider is displayed to the right of the scale only when the subject stops using the mouse. We measure a subject's performance as the number of correctly positioned sliders within the alotted time. Gill and Prowse demonstrated that this measure corresponds closely to the effort exerted by a subject. Before the start of Period 1, subjects were given one minute to practice the task.

We measure creative performance via the "Unusual Uses Task." Originally developed as Guilford's Alternative Uses Task (Guilford, 1967), it was later incorporated in the *Torrance Test of Creative Thinking* (Torrance, 1968, 1998), the most widely used and validated test to assess creativity (Kim, 2006). In our implementation, we follow closely the procedures that psychologists have developed over many decades. In the Unusual Uses Task, participants are asked to name as many, unique and unusual uses for an ordinary item, such as a tin can, as they can. This captures a central element in applied business innovations: the recombination of existing bits of knowledge in novel ways (Weitzman, 1998; Woodman et al., 1993; Simonton, 2004). Specifically, the task requires divergent thinking or "thinking outside the box," which is one of the most important components of the creative process (Runco, 1991). One advantage of the Unusal Uses task for our purposes is that it provides a clean numerical measure of creative productivity.

In the experiment, subjects had to sequentially brainstorm unusual uses for three different items: a sheet of paper, a tin can, and a cord. Subjects were informed that they should not limit themselves to a particular size of the item. Moreover, the unusual use that they come up with could require more than one of the items; for instance, the use could require more than one sheet of paper or several tin cans. The order in which subjects had to work on the items was fixed, regardless of treatment: (1) paper, (2) tin can, (3) cord.

In the creative task, just as in the slider task, subjects had a test period of one minute. In this test period, subjects were given the item "old tire" to familiarize themselves with the task and the input mask on the screen. We used the three standard measures of the Unusual Uses task to evaluate subjects' responses: fluency, flexibility, and originality (Guilford, 1959), and we told subjects how their answers would be scored.[7] Fluency refers to the number of valid answers. An answer is valid if the stated use is possible to implement and the realization is at least vaguely conceivable. Fantastic or impossible uses are not counted. Examples of a valid use of a tin can are, for instance, a flower pot, a pen container, and a drum. In contrast, examples of invalid answers are the use of a tin can as a television, a computer, or a window.[8] In the experiment, each valid use was given one point.

The second evaluation measure, flexibility, reflects the variety of a subject's responses and is determined by counting the number of different categories into which responses fall. For instance, the answer candleholder falls into the category 'decoration', and answers like a rattle or a drum into the category 'musical instruments.' Subjects received one point for each category. Common categories for the tin can include 'non-food containers' (for instance, a pen container), 'sporting goods' (for instance, a football), and 'communication' (for instance, a tin can phone). Overall, there were roughly 55 categories for each of the three items that we used.

Finally, the originality of responses was measured by the statistical infrequency of answers. In order to get an idea of the frequency of responses, we conducted a pre-test with 127 participants who worked on the three items under a fixed wage scheme. We then tabulated all valid answers for each item according to the frequency with which the answer was given and constructed a rating scale to assess answers in the experiment. This scale allotted one additional point to a valid answer if less than 8% ("original") and two additional points if less than 1% ("very original") of participants gave that answer. In comparison to other measures of creativity that rely on expert ratings, our statistical approach to originality is more reliable and objective.[9]

[7]The original Guilford Test uses a fourth criterion for scoring, elaboration, which refers to the degree of detail of the answers. We refrained from using this fourth dimension because it is largely effort-based and would have constrained our capacity to score answers within the time frame of the experiment.

[8]Usual uses, such as a food container in the case of the tin can, were not scored in the original version of Guilford's Alternative Uses Task or the TTCT. However, the original instructions of the test (as well as our instructions) do not explicitly exclude usual uses from scoring. We therefore scored usual uses as valid answers. Excluding usual uses, however, does not alter any of the results reported below (results available upon request).

[9]The answers gathered during the experiment itself allowed us to update the rating scale to more accurately

Examples for an original use of a tin can are an insect trap or an animal house. Very original answers include using the tin can as a scarecrow, a shower head, a treasure chest, or a grill (by putting coal into it and meat on top). Table 1 illustrates further examples of frequent answers and categories as well as original, very original, and invalid responses for all items.

Scoring was conducted by research assistants who were well versed with the scoring procedures and blind to the treatments. The point system that we use is the same as the one used in the Torrence Test of Creative Thinking. To economists, this point system will appear somewhat arbitrary. We will therefore report the results of each evaluation measure separately as well.

It is important to note that while the Unusual Uses task has many desirable attributes for the purposes of our study, it is not without limitations from an economist's perspective. In particular, it captures a particular kind of creative process, "divergent thinking" and it is unclear whether our findings carry over to tasks that require "convergent" thinking, that is, tasks that require finding one unique solution to a problem.[10] One could, in fact, argue that the task is a complex task and that there is no guarantee that a high score in the task actually reflects true creativity in every case (e.g., Kim, 2006). Also, the task measures and rewards both quantity and originality. In many real-world settings, the principal cares about the "best" idea and not about many, original ideas. To the extend that one wants to learn about such situations, one can consider creativity as one input to developing the "best" idea, and this paper sheds light into how this input responds to the provision of explicit and implicit performance incentives. Finally, in order to compensate agents, we used a particular formula that gave numeric weights to the different dimensions. This formula is taken directly from the Torrence Test of Creative Thinking, but there is no particular reason why those weights should be the "correct" ones from the perspective of the principal. In other settings, this metric might look differently, might be less precise, and have different weights. For simplicity, we report our main results in terms of this numeric score, but we also report the performance in the individual dimensions.

3.2 Basic Set up

The experiment uses a principal-agent set up, where subjects are randomly assigned to the role of a principal ("employer") or an agent ("employee"). This feature is important for reciprocity

reflect overall statistical infrequency. The rating scale used for analyzing our results is based on more than 700 subjects (pre-test subjects as well as subjects in all main treatments including the supplementary *Feedback* treatment.) The results do not depend on whether or not we use an updated or the original, pre-test rating scale. Finally, our results are robust to using an expert panel for grading originality rather than the statistical approach (details below).

[10] A prime example of divergent thinking are brainstorming sessions where one tries to come up with has many possible solutions to a problem as possible. Convergent thinking, by comparison, refers to situation in which there is only one or a very limited set of possible solutions and one tries to find it/them. Examples include algorythms or chemical substances with certain properties. Kim (2006) for a brief discussion of the Torrence Test and its relationship to convergent and divergent thinking.

considerations as it allows 1) for voluntary financial transfers from the principal to the agent, and 2) agents' effort to affect the principal's payoff, which gives agents a clear way of reciprocating if they wish to do so. At the start of the experiment, subjects were assigned to groups of five participants, each consisting of one principal and four agents. The role and group assignment remained fixed throughout the experiment.

All sessions were identical in their basic structure: employees were asked to work for the principal for three 3-minute periods on either the routine or the creative task. Treatments (see below) were introduced before the second working period. Figure XXX depicts this basic set up. The three-period design allows us to measure agents' baseline performance under a fixed wage in Period 1, the performance response to the reward in Period 2 – our main focus of interest – and post-treatment performance under a fixed wage in Period 3 – a joint measure of any post-treatment effects related to, for example, treatment specific learning or lasting changes in intrinsic motivation. In order to keep the discussion of the results short and concise, we have relegated the discussion of period 3 effects to the Appendix.

After the three working periods, agents completed a few brief decision tasks as well as questions about their socio-demographic characteristics such as gender, field of study, level of education, high school grade and leisure activities as well as questions regarding their personality traits.

Employers' payoffs consisted of a fixed pay component and a variable pay component that was determined by the performance of the four agents in their group. All payoffs during the experiment were stated in "Taler," the experimental currency unit. The exchange rate was 100 Taler = 1 Euro. In the routine task, principals received 5 Taler for each slider that was correctly positioned by their four agents. In the creative task, principals received 5 Taler for each validity point (valid answer), 5 Taler for each flexibility point (category mentioned), and 5 Taler for each originality point (5 Taler per original answer and 10 Taler per very original answer) given by their four agents. Agents learned about the scoring procedures and the principals' payoff function in the instructions.

In order to create an environment that carried an opportunity cost of working, we offered agents a *time-out button* (Mohnen et al., 2008) which was displayed at the bottom of the screen during all working periods. Each time an agent clicked the time-out button, the computer screen was locked for 20 seconds, prohibiting the entry of creative ideas or the movement of sliders, and 5 Taler were added to the agent's payoff. This procedure has been used in a variety of experiments to ensure that experimental subjects do not merely work on the experimental tasks out of boredom due to the absence of alternative activities (Eckartz et al., 2012; Mohnen et al., 2008). Subjects could push the time-out button as long as the remaining time in the working period was at least 20 seconds. In order to ensure that subjects were aware of the time-out button and understood its usage, we had a trial period that lasted 60 seconds in which subjects could test the time-out button. While the time-out button prevents any "production" in the simple,

routine task, we cannot rule out that subjects in the creative tasks continued thinking about the problem during the time that their screen was locked. This does not mean, however, that subjects in the creative task did not face any opportunity costs of time. The timeout button may still hinder production since it precludes a critical task – entering ideas. Overall, the use of the time-out button was limited, suggesting that research subjects felt their time was better spent completing the assigned task. XXXXX additional info: how often it was used etc etc XXX

3.3 Design and Implementation of Treatments

In order to address our research questions, we implemented a 2 x 3 design consisting of a *Control* group, a *Tournament* treatment, and a *Gift* treatment for both the routine and the creative task. We used a performance tournament rather than a piece rate as the performance-dependent reward scheme because tournaments are widely used in practice to reward individuals for creative performance and innovations (Brunt et al., 2012; Kremer and Williams, 2010). For instance, companies increasingly allocate creative tasks to online platforms with creative contests (such as www.innocentive.com or www.jovoto.com) to complement their in-house research and development. These platforms offer tournament-based compensation for various creative tasks such as scientific problem-solving, software development, and graphic art design (Boudreau et al., 2011).

In the *Control* group, agents were paid a fixed wage in each period and principals were not able to implement rewards. The control group (one per task) allows us to account for learning and fatigue, and to standardize performance across the two tasks (so that we can compare effect sizes).

In the *Tournament* and the *Gift* treatments, principals and agents were informed at the end of period 1 that the principal could invest in a reward scheme for period 2. Regardless of whether the principal decided to implement the reward scheme, agents received information on the type of the reward (tournament or gift, depending on the treatment) and on the associated costs to the principal. Before the start of Period 2, agents learned whether their principal had instituted the reward. Subjects were also told that principals did not receive any information about their agents' performance until the very end of the experiment. This ensured that agents perceived the wage gift as "kind," rather than as compensation for good performance in Period 1. Moreover, it avoided an endogenous selection of rewarded agents based on Period 1 performance.

In both treatment groups, principals and agents received a fixed wage of 300 Taler at the beginning of each period. In the *Gift* treatment, the principal had to decide whether or not to provide an additional monetary gift of 300 Taler to each of her four agents at a total cost of 200 Taler to herself.[11] In the *Tournament* treatment, the principal could also transfer a

[11]The use of efficiency factors is common practice in the experimental literature on gift exchange (see, for

total of 1200 Taler to her four agents at a cost of 200 Taler to herself. This feature allows us to draw conclusions about which of the two reward schemes is prefereable given a fixed budget for rewards. However, the payment structure was different in the tournament than in the gift treatment. In the tournamnet treatment, agents' performance dictated whether or not they received a reward. Specifically, the top 50% of performers (two out of four agents) received a bonus of 600 Taler each in Period 2, whereas the bottom 50% received nothing. Subjects in the *Control* group also received a fixed wage of 300 Taler in Periods 1 and 3, and a fixed wage of 600 Taler in Period 2. This increase in Period 2 mirrors (expected) payoffs in the treatment groups. This ensures that any performance difference between the treatment and the control groups is driven by the treatments rather than by endowment effects.[12] In essence, therefore, the two treatment groups and the control group are budget equivalent ways of compensating workers and the results will show which way of compensating workers results in the highest output.

When learning about the principal's reward decision, employees also learned that performance would be evaluated immediately after Period 2 ended, and that the winners and losers of the *Tournament* would be revealed before Period 3 started. Finally, after Period 2 and after the revelation of winners and losers (presented as private information on a subject's screen), subjects in both treatments were informed that there would be no further rewards. In all treatments (including the control group) it was further announced that the payment structure in Period 3 would be identical to that in Period 1.[13]

This study focuses on agents' responses to rewards, rather than on principals' reward decisions. Therefore, the rewards were relatively cheap for the principal, and endowments in the control group mirrored (expected) payments in the two treatment groups when the principal opts for the reward. Specifically, endowments in Periods 1 and 3, periods without rewards, are identical in the control and the two treatments groups: the principal as well as each of her four agents

instance, Brandts and Charness, 2004) and is thought of as representing situations in which gifts are more valuable to the recipient than to the donor. The attractiveness of the reward was important in our setting because we were mainly interested in agents' responses to rewards rather than in whether or not principals opted for the rewards.

[12]It is possible that the higher pay in the control group in period 2 might induce a reciprocal response (towards the experimenter). A priori it is unclear, however, whether reciprocity should increase or decrease performance (increased performance might be considered desirable but leads to additional expenses for the experimenter because he needs to pay the variable pay component to the principal). Hennig-Schmitt et al. (2010) documents that wage gifts are not reciprocated when the payoff function of the experimenter is unknown to subjects, which is the case here. Similarly, the institution of the tournament could induce a reciprocal response (towards the principal). To the extent that this is true, the tournament effect – just like any tournament effect in the literature – comprises both the incentive effect of the tournament and the reciprocity that goes hand in hand with it.

[13]The tournament treatment effect is a reduced form that captures a range of possible channels via which worker output is affected. For one, tournaments give workers strong incentives, which may increase effort. Also, tournaments might be perceived as a gift from the principal (it does cost the principal money). To the extent that this is true, the tournament treatment effect also reflects the combination of incentives, reciprocity, and any other factor that might be at work.

receive 300 Taler.[14] In Period 2, principals in the control group receive an endowment of 100 Taler and agents an endowment of 600 Taler each. This mirrors the expected payoffs in the treatment groups when the principal opts for the reward scheme (unconditional gift or tournament, depending on the treatment).[15] This procedure ensures that any performance differences between the treatments and the control goup are solely driven by the rewards and not by other factors such as distributional concerns or income effects.[16] Table 6 in the Appendix provides an overview of the fixed and variable pay components for all periods, treatments, and roles.

3.4 Procedures

The experiment was conducted at the experimental laboratories at the universities of Frankfurt, Mannheim, and Heidelberg, Germany. Participants were recruited via the Online Recruitment System ORSEE (Greiner, 2004). The experiment was computerized using the software z-Tree (Fischbacher, 1999).

All interactions within the experiment were anonymous and communication was not allowed. Subjects were seated randomly at a computer workstation upon arrival and were provided with hard copy instructions that detailed the random matching of groups and roles ("employer" or "employee"), the basic structure of the experiment, the task (routine or creative), and the scoring procedures. A translation of the original instructions can be found in the Appendix. A few pieces of information, such as fixed wages in Periods 2 and 3 as well as the availability of rewards and reward decisions, were presented on the computer screen during the experiment. Before the experiment started, subjects had to complete a series of questions about how their actions would determine their own and their principal's payoffs to ensure that they understood the instructions.

At the end of the experiment, subjects' payoffs in the experimental currency unit "Taler" were converted into Euros at an exchange rate of 100 Taler = 1 Euro. Subjects were paid in private. Sessions (including instructions, the experiment, and questionnaires) lasted about 75 minutes and subjects earned 15 Euros on average.

[14] On top of their endowment, principals earn additional money from the performance of their agents.

[15] If the principal decided against the reward, the principal and her agents received 300 Taler each as fixed Period 2 endowments in both the *Gift* and *Tournament* treatments (identical to payments in Periods 1 and 3), while they earned 100 and 600 Taler in the control group, respectively. Therefore, we cannot assess responses to negative reward decisions in an experimentally clean way.

[16] See, e.g., Fehr and Schmidt (1999) on inequality aversion. Our set up also disentangles pure intention-based reciprocity from other distributional concerns (Charness, 2004) by allocating the same payoffs to subjects in the treatment groups (with positive reward decisions) and the control group in Period 2. The only difference that remains is that control group payoffs were exogenously imposed by the experimenter, while treatment payoffs were chosen by experimental subjects, the principals.

4 Main Results

4.1 Descriptive Statistics

Table 8 provides an overview of the descriptive statistics split up by treatment and task. XXX ADJUST NUMBERS GIVEN THAT WE CHANGEd TABLE XXX [17] As we are interested in assessing how rewards affect performance, we ran sessions until each treatment included roughly 60 agents with a positive reward decision. Overall, we observe 116 rewarded employees and 60 control group employees working on the slider task and 116 rewarded employees and 56 control group employees working on the creative task in our main treatments. The treatments are largely balanced with regard to the location of the experiment, gender, age, and field of study, albeit some differences in mean age of participants are statistically significant. As we will show in our main analyses, controlling for these characteristics does not alter the results.

The last row of Table 8 displays means and standard deviations of the baseline performance in Period 1 across treatments and tasks. Mean performance varies between 16.6 and 22 in the slider task, where performance is measured by the number of correctly positioned sliders within each three minute work period. In the creative task, average performance varies between 16 and 18. Performance here represents a subject's score in the creative task (see Section 3 for details on the scoring procedure). Apart from the *Tournament* treatment in the slider task, there are no statistically significant differences between the treatments and the respective control groups in either task. In the former, individuals' performance is slightly better in Period 1 than that of subject in the control group (Wilcoxon rank-sum test, $p<0.05$). To account for these initial performance differences, we control for baseline performance in the analyses that follow and use the change in performance between Periods 1 and 2 as the outcome variable. XXXX HERE ADDITIONAL INFO FROM REF REPORT ANSWERSXXX

4.2 Results

Figure 3 displays the change in *raw* performance from Period 1 to Period 2 by treatment and task. The figure shows that performance in the slider task increases across both treatment groups between the first and second round. The performance increase is particularly strong in the *Tournament* treatment, but is also clearly detectible in the *Gift* treatment. The moderate performance increase in the control group is consistent with the learning effects documented

[17] Our full balance table can be found in the Online Appendix. The full balance table contains obversations from a host of supplementary treatments that we ran, totalling 1123 subjects. An additional 9 employees were excluded from the analysis due to an insufficient knowledge of the German language.

in prior studies using this task (e.g., Gill and Prowse (2012) and Araujo et al. (2016)).[18] The pattern is somewhat different in the creative task where only agents in the *Tournament* treatment improve their performance between Periods 1 and 2. There are no notable changes in mean performance in either the *Gift* treatment or the *Control* group.

Thus, the raw data suggest that performance increases in the creative as well as in the simple task in response to the tournament. However, only employees in the simple task seem to respond to the gift.

To control for possible confounding factors, we fit the following regression model using ordinary least squares (OLS). We make performance comparable between the different periods (and thus different items on the creative task) and the two different tasks by standardizing performance.[19]

The full model takes the form:

$$\begin{aligned}\text{Std. Performance Period 2}_i = \beta_0 &+ \beta_1 \text{ Std. Performance Period 1}_i \\ &+ \beta_2 \text{ Std. Performance Period 1}_i \times \mathbb{1}_{\text{ Slider-Task }_i} \\ &+ \beta_3 \text{ Gift}_i + \beta_4 \text{ Gift}_i \times \mathbb{1}_{\text{ Slider-Task }_i} \\ &+ \beta_5 \text{ Tournament}_i + \beta_6 \text{ Tournament}_i \times \mathbb{1}_{\text{ Slider-Task }_i} \\ &+ \gamma X_i + \epsilon_i.\end{aligned} \tag{1}$$

Standardized Period 2 performance of individual *i* is the dependent variable. It is regressed on *i*'s baseline performance in Period 1 as well as on the treatment dummies, and, in the most comprehensive model, a set of person-specific control variables (X_i). Treatment dummies as well as baseline performance are interacted with a dummy that indicates the type of task (here: the simple task). This allows the treatment effects as well as the impact of baseline performance to differ between the creative and the simple task. The creative task control group serves as the reference category. Standard errors are adjusted for potential heteroscedasticity in all regressions.[20] Note that the main regressions only include individuals where the principal opted for the tournament or the gift; Individuals with negative reward decisions are excluded from these analyses. We analyze treatment effects using interaction effects as this allows us to directly display our two outcomes of interest: treatment effects on creative performance (the main treatment effect) as well as an indicator of whether treatment effects differ between the simple and the creative task (the interaction effect). The treatment effect on the simple task is

[18] In our setting, performance could also increase because of the higher endowment in Period 2.

[19] We use the standard approach of subtracting the mean performance of the control group (in the respective working period and task) and dividing the resulting difference by the standard deviation of the control group in the respective task. Therefore, the standardized performance of the control group has a mean of zero and a standard deviation of one. Treatment dummies (and additional controls) remain non-standardized to ease interpretation.

[20] The results are robust to using cluster-robust standard errors (by session) that control for potential intra-session correlation.

equal to the sum of the main treatment effect and the interaction effect.

The results are shown in Table 3. Column I presents the most parsimonious specification and does not control for baseline performance. Column II of Table 3 adds controls for baseline performance. Column III adds additional control variables including age, age squared, sex, field of study, time period (semester, exam period, semester break), and location. We also ran more comprehensive specifications with further control variables, such as the Big Five and incentivized risk and reciprocity measures, but their inclusion does neither improve the fit of the model nor does it affect the results. As can be seen, the results are robust to the inclusion or exclusion of all of our additional control variables. When discussing the results below we refer to the full specification in Column III.

XXXCHECK WHETHER NUMBERS ARE STILL CORRECT GIVEN THAT WE CUT FEEDBAACK TREATMENT OUT OF TABLEXXX Consistent with the raw data shown earlier, the *Tournament* treatment has a large and statistically significant positive effect on performance in both tasks. The coefficient on the *Tournament* treatment represents the impact of the tournament on performance in the creative task. The interaction effect between the *Tournament* treatment and the slider task suggests that there is no statistically significant difference between the impact of the tournament on the creative task and the slider task. This suggests that creative performance is responsive to incentives and that tournaments have the power to substantially increase creative output. Interestingly, the *Tournament* effect is very similar in both tasks as can be seen by the interaction effect: Agents increase their performance by approximately 0.7 standard deviations as compared to the control group in both the creative and the simple task.[21]

While the *Tournament* worked well in stimulating creative performance, the data suggest that the *Gift* was less effective. The effect of the gift is substantially smaller than that of the tournament for both tasks (Wald test, $p<0.01$ for both tasks). In the simple task the *Gift* induced an economically and statistically significant effect ($p<0.05$). The effect of the gift in the simple task is represented by the main gift treatment effect and the interaction effect of the gift with the slider task. The effect size of roughly 0.2 standard deviations is well in line with what other studies on reciprocity have found.[22] In contrast, the point estimate of the impact of the *Gift* treatment in the creative task is close to zero across specifications.[23] The standard error on this

[21] We ran an additional *Feedback* treatment to examine to what extent the tournament treatment effect was driven by the money and to what extent it was driven by concerns for a good relative performance rank. The *Feedback* treatment mirrors the *Tournament* treatment in that it reveals who is and who is not a top 2 performer but in contrast to the *Tournament* treatment there are no financial consequences to being a top 2 performer. This treatment raises performance by roughly a third of what we observe in the tournament treatment in both tasks. Results of this treatment in periods 2 and 3 can be found in Table XXX in the Webappendix.

[22] We find no evidence for gender differences in either the *Tournament* or the *Gift* treatment.

[23] Interestingly, it appears as though at least some of the principals might have anticipated that the gift would not work well in the creative task. In the *Gift* treatment only roughly half of the principals opted for the gift in the

point estimate is roughly 0.1. A key question is how to interpret this zero point estimate and its associated non-trivial confidence interval. One potentially interesting hypothesis test is a test of the equality between the effect of the gift on the simple task and on the creative task. We reject this hypothesis at the 0.1 level, i.e. there is evidence against this equality but it is weak. But in light of the previous research which has repeatedly produced a 0.2 standard deviation impact of gifts on performance in the simple task, another relevant null hypothesis to test is whether the point estimate of the gift on the creative task is significantly different from 0.2. This hypothesis we reject comfortably at the 0.05 level. In essence, we can say with substantial confidence that the effect of the gift on the creative task is smaller than the 0.2 that is generally found on the simple task. But we cannot with confidence conclude that the true effect of the gift on the creative task is zero even though the point estimate is zero. [24]

The analysis thus far has focussed on positive reciprocity. The structure of our experiment also provides information with regar to negative reciprocity. Agents are informed if the principal chooses not to institute a gift or tournament, which has the potential for inducing negative reciprocity (both the gift and the tournament involve a transfer of an additional 300 Taler from the principal to the agents). If negative reciprocity is present, then among the agents who do not receive the gift or tournament, output should decline relative to the respective control group. Because negative reward decisions are relatively rare (a total of 116 agents) our results in this domain have to be interpreted with care.[25]

Figure 4 shows the results. The solid bars represent agents' performance responses (as compared to the control group) to positive reward decisions. These effect sizes correspond to the main regression results discussed above. Hatched bars represent agents' performance response to the announcement that the principal had not invested in additional rewards. Again, an asymmetry between the two tasks emerges when looking at the *Gift* treatment. Agents reduce their performance sharply in response to a negative *Gift* decision in the slider task, suggesting negative reciprocity. By comparison, they remain relatively unaffected in the creative task. These results are the mirror image of those discussed above, providing support for the conjecture that reciprocity is much stronger in the simple task than in the creative task.[26]

---

creative task compared to three quarters of the principals in the slider task. Three quarters of the principals in the *Tournament* treatment invested in instituting the tournament. This is true in both the simple and the creative task.

[24] Since standard errors go down with the square root of sample size, the standard error would still be 0.07 even if we doubled the sample size. This would, of course, increase the statistical significance of any hypothesis test, but it would take radical increases in sample size (maybe ten times the sample size) to generate estimates so precise that one would talk of a true zero effect with confidence.

[25] Also, the fixed payments in the control group were designed to be equal to (expected) payments under positive, but not negative reward decisions. Hence, treatment and control groups are not payoff equivalent in cases of negative bonus decisions. Principals and agents earn a fixed wage of 300 Taler in Period 2 when the principal decides against giving a reward, whereas they are endowed with 100 Taler and 600 Taler, respectively, in the control group.

[26] Two other potentially interesting observations can be seen in Figure 4. First, in the slider task the response to a negative reward decision is substantially larger than the response to a positive reward decision. This is in line

There two channels via which workers can increase productivity. One is reduced usage of the time-out button, the other is greater effort during actual work time.

They hit the time-out button more or less XXX in this or that treatment.... XXX

productivity of minute or work as compared to minute of treatment... XXXX

We can look at the use of the time-out button to examine whether the effects in the *Tournament* and the *Gift* treatment are driven by a reduction in the number of breaks or by a performance improvement during actual work time. Table XXX shows the results. XXXXXX In both tasks, the *Tournament* effect remains positive and significant (albeit much smaller) after controlling for the number of breaks. In contrast, the performance increase in response to the *Gift* in the slider task seems to be solely driven by a reduction in breaks and, hence, by an extension of actual working time. XXX MORE CONTENT HERE

4.3 Sensitiviy Analysis

Thus far, we have reported results fro the creative task in terms of one score that captures three different dimensions of creativity: 1) the number of valid uses (validity), a measure of quantity, 2) the number of categories into which answers fall (flexibility), and 3) the originality of ideas (statistical infrequency of a response).

Table 5 shows the results split up by these three different dimensions. The pattern of results that we discussed for the overall score in our main analyses is reflected in all three dimensions of creativity. This result is, however, somewhat mechanical because each valid idea generates one point for certain in the validity dimension and is eligible for more points in flexibility and originality. This leads to a strong positive correlation across the three dimensions of creativity. In order to assess the degree of flexibility and originality conditional on the total number of valid ideas, we created measures which we denote the "flexibility rate" and "originality rate." For both of these variables we divide the number of flexibility/originality points of a subject by the number of valid answers given. Columns V and VI display the results. Conditional on the number of uses, neither treatment has a statistically discernable effect on idea flexibility. The tournament does, however, increase originality rate. That is, XXXXX. This is evidence against the notion that the strong performance increase in the *Tournament* treatment came at the expense of idea originality. The opposite is true.

MAYBE THIS BE ITS OWN COLUMN..[27]

with other evidence showing that negative reciprocity tends to be stronger than positive reciprocity (Kube et al., 2013). Second, in neither task does the withholding of a tournament reduce performance. One explanation of this is that in response to a tournament agents put out a lot of additional effort. In that sense, tournaments might not be viewed as gifts to agents even though they cost the principal resources.

[27] One might be concerned that statistical infrequency does not accurately reflect what is commonly understood as an "orignal" idea. Therefore, we checked the robustness of our results on originality using subjective originality

Our set-up mirrors settings in which the principal cares about as many and as original ideas as possible. In other settings (architecture competitions, for example), the principal cares about identifying the "best" idea/project. To address this issue, we tried to map our data more directly into what might go into a principalâĂŹs payoff function in these settings. In particular, we went back to the data and had a research assistant - blind to treatment - code any answers that they considered a "best" answer in a subjective and unspecified way. Overall, X answers (or roughly Y% of all ideas) were coded as "best." Even though subjects were not incentivized to produce "best" answers, we were curious to see if there were treatment effects in this dimension. The results are displayed in Column VII of Table Y. The results show that XXX. Hence, the tournament incentive not only increased performance along the dimensions that subjects were incentivized, but also increased the number of "best" ideas. That is, incentivizing agents to produce as many and as original ideas as possible also increased the number of âĂIJbestâĂİ ideas. Obviously, this does not guarantee that the *Tournament* treatment would have worked well, had subjects been directly incentivizied to provide a "best" idea.

One might also wonder if treatments affected the number of invalid uses. Depending on the treatment, there were, on average, between 1 and 1.5 invalid answers per person in period 1. That is, between 10 and 17 percent of answers were classified as invalid in period 1. Column VIII shows that there is no statistically significant treatment effect of either the *Tournament* or *Gift* on the frequency of invalid uses.

Taken together the different pieces of evidence discussed above suggest that there is a difference in subjects' responses to the gift across the two tasks and that the gift was less effective in the creative task than in the simple task.

One explanation for less reciprocity in the creative task would be that agents are simply unable to increase output even though they try to. The results of the *Tournament* treatment speak against this explanation. In response to tournaments agents sharply increase their output showing that they can do so if they wish.

Another possibility is that agents did not want reciprocate. The results from the simple task speak against this notion. Agents in the *Gift* treatment in the slider task receive a gift of the same magnitude and exhibit a reciprocal response. While we do not have a particular theory as to why the desire to reciprociate may differ across the two tasks, we nonetheless test this hypothesis in the supplementary treatment described below.

A third possibility is that they did want to reciprocate but that the nature of the creative task

ratings. Towards that end, we asked five research assistants unfamiliar with the experiment and blind to the treatments to evaluate the originality of each answer. The evaluators were instructed to assign one point to answers that they perceived as original, two points to answers that they perceived as very original, and zero points otherwise. Using the score from this subjective originality assessment as the dependent variable does not alter the results: originality increases under the *Tournament* incentive. The *Gift* treatment, by comparison, even lowers originality by this measure. Results available from the authors upon request.

makes it hard for them to do so. This would be the case, for example, if agents get utility from providing appropriately sized reciprocal responses, but disutility from reciprocting in a manner that is either too small or too large (relative to the size of the gift).[28] Unlike in the slider task where agents have perfect control over how many sliders they move and, hence, how much benefit they generate for the principal, this is more opaque in the creative task. Although the procedure for evaluating ideas was detailed in the instructions, subjects could not perfectly predict the exact number of points that they were generating for the principal because agents' scores depended on ratings (validity, flexibility, and originality). It is therefore harder for agents to target an appropriately sized gift in the creative task than it is to do so in the simple task. In particular, following the literature on inequality aversion (e.g., Fehr and Schmidt, 1999) one might hypothesize that it is more costly in utility terms for an agent to reciprocate too much than it is to reciprocate too little (relative to the gift). Under these conjectures, a fear of regret of disproportionally large reciprocity may lead agents not to exert effort in the gift treatment in the creative task. Hennig-Schmitt et al. (2010) and Englmaier and Leider (2010) provide initial evidence that uncertainty about how own effort affects the principal's surplus can reduce agents' reciprocal behavior. This mechanism should be more relevant in the *Gift* treatment where agents have no financial stake (fixed wage) and their effort is solely determined by a mix of intrinsic motivation, ability and goodwill towards the principal, and should be less relevant in the *Tournament* treatment where agents most likely trade-off the cost of effort with the perceived likelihood of winning, which is independent of the profit they generate for the principal.

To test both of these hypotheses (lack of desire to reciprocate in the creative task; inability to finetune reciprocal response), we ran two supplementary treatments: *Creative Transfer Control* and *Creative Transfer Gift*. These treatments were identical to the control group and the *Gift* treatment in the creative task apart from the following aspects: First, at the end of each period, agents were told how many Taler they had generated in the preceding round. Second, they could then decide how much of this surplus to transfer to the principal. Hence, agents had perfect control over how much money would accrue to the principal as a result of their work. Surplus that was not transferred was "lost" in the sense that it did not benefit the principal (and agents' payments were independent of their performance). As in the *Gift* treatment, the principal had the option to provide a monetary gift of 300 Taler to each of her agents before the start of Period 2 in the *Creative Transfer Gift* treatment. The treatment *Creative Transfer Control* serves as an additional control group that allows us to seperate out a change in the response to the gift from changes in behavior that are solely driven by either the provision of feedback on the number of points earned or by the ability to choose a transfer amount. Comparing behavior in *Creative Transfer Gift* with that in *Creative Transfer Control* therefore allows us to assess whether reciprocity emerges when agents have the power to fine-tune their back transfers in the

[28] See, for example, Dorsch and Kelley (1994) on the relationship between the size of a gift and the resulting sense of psychological indebtedness.

creative task.

Table 8 shows that baseline performance is very similar in *Creative Transfer Control* and *Creative Transfer Gift* and slightly higher than average performance in the other treatments utilizing the creative task.[29] Only about two thirds of all subjects transfer the maximum amount in Period 1. About 20% transfer less than half of the generated surplus while the remaining subjects transfer intermediate amounts. In the following analysis, we will focus on transfers, as transfers and not actual performance influence principals' payoffs and thus signal reciprocity. As in the main analysis, we control for baseline transfer – each agent's transfer in Period 1 – in all regressions.

The results are striking. Agents in the *Creative Transfer Gift* treatment transfer significantly more of their surplus in Period 2 (after the gift) than in Period 1 (Wilcoxon signed-rank test, $p<0.01$). They also transfer significantly more of their surplus in Period 2 than do agents in the *Creative Transfer Control* group in the same period (Wilcoxon rank-sum test, $p<0.05$). Figure 5 depicts treatment effects for all *Gift* treatments. The bars show coefficients from separate OLS regressions analyzing the effect of the *Gift* treatment on Period 2 effort (or, in the case of the supplementary treatments, amount transferred), controlling for baseline performance (transfer). The graph shows not only that subjects reciprocate the gift in the *Creative Transfer Gift* treatment, but also that the order of magnitude of the effect is similar to the increase in output in the *Gift* treatment in the simple task. Column IV of Table 3 reports the results from the associated regressions. Again, one can see that reciprocity emerges in the *Creative Transfer Gift* treatment and that the effect is similar to that in the simple, routine task.[30][31]

Taken together, the findings from these two additional treatments suggest that the lack of reciprocity in the creative task was driven solely by agents' inability to control and fine-tune their impact on the principal's bottom line. When agents have that control, they reciprocate the

[29]This may not be surprising as the transfer option allows agents to decouple their work from what accrues to the principal. Therefore, agents who intrinsically enjoy working on the task but that are also inequality averse, can work harder in these treatments without fearing that they create unduly large profits for the principal.

[30]Performance itself is not affected by the gift, which is not surprising given that many subjects had substantial leeway in increasing transfers without increasing performance.
Just like in the main *Gift* treatment, we find no evidence for gender differences in response to the gift in the *Creative Transfer Gift* treatment.
With respect to principals' reward decisions, only 8 participants were affected by a negative reward decision by the principal, making it impossible to reliably estimate this effect. The existing data, however, suggest that negative reciprocity re-emerges in the *Creative Transfer Gift* treatment with a mean performance decrease of 0.82 in response to a negative gift decision, similar to what was observed in the slider task.

[31]Interestingly, this asymmetry (agents in the creative task have uncertainty over the value of their work to the principal, while those in the simple task do not), did not affect effort in the tournament treatment, where we see similar effect sizes in both tasks. This is probably owed to the fact that agents had a financial stake in the game in the tournament treatment – the tournament prize. In the gift treatment, by comparison, agents work for a fixed wage and their efforts solely affected the profits of the principal. Hence, we conjecture that agents were more concerned about the impact of their efforts on the principal in the gift treatment than they were in the tournament treatment. XXX this footnote into discussion??

wage gift in both tasks and the effect sizes are of similar magnitude. Such control is, however, typically relatively low for creative tasks as well as for other complex problem solving tasks that are prevalent in many white-collar jobs. In these settings information on the exact value of effort (and ideas) to the principal is often not available. Thus, in practice this specific feature inherent in creative tasks may hinder the emergence of reciprocity.

5 Conclusion

XXXhighlight that task rewards both quantity and originalizty dimension and that both contribute to the score in the task XXX talk about how imortant originality is for understanding results.. is this paper really about routine versus creative or about routine versus any type of non-routine. XXX point that many and original ideas might not necessarily be "best" idea and in many settings principal may only be interested in the best idea... nevertheless, many original ideas might be a great input for development of best idea. from ref report... use almost word for word: Yet in most real-world settings, including examples like the ones cited in the paper (e.g., innovation inducement prizes, software coding contests, architecture competitions, etc.), what matters is product quality or performance, not originality, and quality is typically what gets rewarded: the agent with the âĂIJbestâĂİ idea wins, not the agent with the âĂIJmostâĂİ or âĂIJmost originalâĂİ ideas. Creativity may be a necessary input to developing a good idea, but itâĂŹs not an end in and of itself.

XXX insert something that says that it is surprsing that reciprocity goes away /is reduced so much!! one would expect reduction but not necessarily eleminiation of reciprocity.

we learn how incentive affect creativity in a task with a known score metric, in real world people also often understand different dimensions on which they will be rewarded; if this is the best point and rewarding scheme is beyond the scope of this paper

- There has been very little research in creativity, .. so anything about generalizability would be pure conjecture

how relevant is setting in which p cares about output of each individual agent rather than only about best idea (examples),

academics is good example where high number of original ideaas at outset of career because we donâĂŹt know for sure what idea might turn out to be valued by market; in graduate class that; important in ccases in which as you come up with ideas one does not know final value. Go after a lot and crazy ideas âĂę .free disposal

is it a problem that more ideas lead to more creative ideas?

This paper reports the results from a large-scale laboratory experiment that studies the impact

of both explicit incentives (tournaments) and implicit rewards (wage gifts) on creativity. To the best of our knowledge, this is the first study to analyze the impact of wage gifts on creativity. We also have not come across another study that compares the effectiveness of wage gifts and tournaments in one set-up, providing insights into the relative effect sizes of these two rewards on both a creative and a simple task. The inclusion of the simple task allows us to benchmark the effectiveness of the different rewards on creativity and to link our results on creativity to the existing literature on tournament incentives and wage gifts that uses simple tasks.

We report two sets of interesting results. The first relates to the tournament incentive. Our results suggest that a tournament prize for above-average performance has a substantial positive incentive effect on creativity. The effect size is similar to that on performance in the simple task. This indicates that incentives can influence creativity and that there is no crowding out of intrinsic motivation.[32] About one fourth of this effect seems to be driven by a concern for relative rank, as is suggested by a supplementary treatment in which subjects work towards receiving the same rank information as in the tournament but without financial consequences. Thus, it is largely the monetary prize that is responsible for the positive incentive effect of the tournament. Interestingly, the tournament does not only increase the quantity of ideas submitted, but also their quality in terms of originality. Further, tournament winners continue to outperform their own baseline performance and agents in the control group. A set of supplementary treatments suggests that this is driven by an increase in task-specific motivation that does not spill over to other tasks. Losers of the tournament show no signs of demotivation in both tasks. The effectiveness of tournaments is in line with the observation that creative tasks are often organized in a tournament framework in the real world. For example, architects on virtually all major projects are chosen via a winner-take-all competition. The same is true on most innovation platforms such as innocentive.com that many companies now utilize for creative input.

A second set of interesting results relates to the financial gift. We find an asymmetry in its effectiveness between the two tasks. While the gift effectively triggers a reciprocal response in the simple, routine task, there is no evidence for reciprocity in the creative task. This suggests that the incentive response function differs between creative and simple tasks, despite the similarity of responses in the *Tournament* treatment. Interestingly, this asymmetry holds for both positive and negative reciprocity, and principals seem to have anticipated this asymmetry as many more principals opted for the gift in the simple task than in the creative task. We explore a set of explanations for this finding and, through the implementation of additional treatments, can trace the effect back to agents' lack of knowledge about how exactly their effort translates into profit for the principal. While agents perfectly observe how much output they

[32] Erat and Gneezy (2015) find evidence for choking under pressure in their creative task that is more "blue skye" in nature.

produce in the simple task, there is some uncertainty in the creative task because the value of their ideas to the principal depends, for example, on an originality rating. This is true for our creative task, but also holds for creative and complex tasks more generally. One implication of our finding therefore is that wage gifts might not boost performance in white-collar jobs that involve complex tasks and creativity. This is important to note as creative and complex tasks are typically governed by incomplete contracts that might have rendered gift exchange a viable way of increasing effort. These results also speak to the ongoing debate about reciprocity in the lab versus in the field (e.g., Kube et al., 2012 or Kessler, 2013) and suggest that one reason for the observed absence of reciprocity in the field could be agents' imperfect knowledge about how their effort affects their principal's profits.

But even if potential benefits from ideas were fully transparent, our study suggests that wage gifts are less effective than tournaments in triggering agents' performance in both tasks. While the tournament was profitable for the principals that opted for it, the gift was not.[33] In this study, tournaments clearly emerge as preferable to wage gifts in terms of fostering creativity. Nevertheless, tournaments do have well-understood downsides that should be considered before implemention. For instance, tournaments have been shown to increase sabotage among workers (for instance, Harbring and Irlenbusch, 2011), to induce a self-selection of more risk-tolerant agents (for instance, Eriksson et al., 2009; Dohmen and Falk, 2011), and to make low performers more likely to give up early in the contest (for instance, Berger et al., 2013).

Of course, our study has many limitations. Studying creativity is difficult. Limitations of the task : While there are Unusual Uses task has many virtues (quantifiable, well-established in creativity reserach, captures multiple dimensions of creativity), it is in the end a particular task - unfamiliar task - it is in a sense contrived (even though it is a proxy for creativity in teh real world) no one outside the lab actually works on this particular task

Creativity is an amorphous concept and there is every reason to think that the particular context or challenge might affect

XXX Liitations of alternative uses task// from discussion with Steve: does not cover process in which creative solution to one problem; not put to market test (dollar value); this is obviously a ver specific task in a somewhat contrived setting relative to , say, in a place where blue skyp creativity is iportnat like in an advertising agency and only onen solution can be presented. But the virtue is that it is a task that is easy to communicate and understand with a straightforward incentive/payoff scheme that taps into the sorts of skills that might be important for understanding creativity outside the lab fact that ratings are not tied to usefulness is a limitation (albeit we go halfway by not counting invalid uses) does nto apply to tasks where

[33]The payoff consequences of the *Gift* in comparison to the control group (difference in average effort per work group minus costs of the gift) are -0.27 Euro in the slider task, -1.82 Euro in the creative *Gift* treatment and -0.29 Euro in the treatment *Creative Transfer Gift*. In the *Tournament*, principals' payoff increased by 1.39 Euro in the slider and 1.45 Euro in the creative task.

quantity does not matter at all does not mirror onein one a single creative task in real world but captures underlying processes of many setting in which ideas can freely be disposed of.

generalizability

section about how tournament m ight be more generalizeable becuse it can also be used when only best idea matters. also talk about where directly applies: brainstorming sessions etc; what we wrote to ref 1

The present study calls for future work that addresses whether the positive tournament effect that we find, also holds for contests with high-stakes. Previous studies indicate that high stakes cause choking under pressure and, hence, a non-linear relationship between reward size and effort, and that this is particularly true for cognitively challenging tasks (Ariely et al., 2009; Bracha and Fershtman, 2012). Further, creative tasks come in many different forms. Charness and Grieco (2012) show that responses to incentives might differ between tasks that have a clearly delineated solution and tasks that are open and require pure out-of-the-box thinking. In our opinion, our task serves as a good proxy for everyday idea generation in firms. However, organizations also rely on other types of creative input that our task could not capture, such as idea implementation or breakthrough innovations. Future work needs to test the robustness of our findings for other kinds of complex and creative tasks.

References

Akerlof, G. A., 1982. Labor contracts as partial gift exchange. Quarterly Journal of Economics 97 (4), 543–69.

Amabile, T., 1996. Creativity in context: Update to "The Social Psychology of Creativity.". Westview Press, New York.

Amabile, T. M., 1997. Motivating creativity in organizations: On doing what you love and loving what you do. California Management Review 40(1), 39–58.

Amabile, T. M., 2013. Encyclopedia of Management Theory. Sage Publications, Ch. Componential Theory of Creativity.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., LamÃl', D., Vesterlund, L., Wang, S. W., Wilson, A. J., 2016. The slider task: an example of restricted inference on incentive effects. Journal of the Economic Science Association 2 (1), 1–12.

Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N., 2009. Large stakes and big mistakes. Review of Economic Studies 76 (2), 451–469.

Autor, D. H., Levy, F., Murnane, R., 2003. The skill content of recent technological change: An empirical exploration. The Quarterly Journal of Economics 118 (4), 1279–1333.

Berger, L., Klassen, K. J., Libby, T., Webb, A., 2013. Complacency and giving up across repeated tournaments: Evidence from the field. Journal of Management Accounting Research 25 (1), 143–167.

Boudreau, K. J., Lacetera, N., Lakhani, K. R., 2011. Incentives and problem uncertainty in innovation contests: An empirical analysis. Management Science 57 (5), 843–863.

Bracha, A., Fershtman, C., 2012. Competitive incentives: Working harder or working smarter? Management Science, 1–11.

Brandts, J., Charness, G., 2004. Do labour market conditions affect gift exchange? Some experimental evidence. The Economic Journal 114 (497), 684–708.

Brunt, L., Lerner, J., Nicholas, T., 2012. Inducement prizes and innovation. The Journal of Industrial Economics 60 (4), 657–696.

Byron, K., Khazanchi, S., 2012. Rewards and creative performance: A meta-analytic test of theoretically derived hypotheses. Psychological Bulletin 138 (4), 809–830.

Charness, G., 2004. Attribution and reciprocity in an experimental labor market. Journal of Labor Economics 22 (3), 665–688.

Charness, G., Grieco, D., 2012. Individual creativity, ex-ante goals and financial incentives, mimeo.

Deci, E. L., Koestner, R., Ryan, R. M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125 (6), 627–668.

Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. The American Economic Review 101 (2), 556–590.

Dorsch, M. J., Kelley, S. W., 1994. An investigation into the intentions of purchasing executives to reciprocate vendor gifts. Journal of the Academy of Marketing Science 22 (4), 315–327.

Dutcher, G. E., 2012. The effects of telecommuting on productivity: An experimental examination. the role of dull and creative tasks. Journal of Economic Behavior & Organization 84, 355–363.

Eckartz, K., Kirchkamp, O., Schunk, D., 2012. How do incentives affect creativity? Working Paper.

Eisenberger, R., Rhoades, L., 2001. Incremental effects of reward on creativity. Journal of personality and social psychology 81 (4), 728–741.

Eisenberger, R., Shanock, L., 2003. Rewards, intrinsic motivation, and creativity: A case study of conceptual and methodological isolation. Creativity Research Journal 15 (2-3), 121–130.

Englmaier, F., Leider, S., 2010. Gift exchange in the lab – it is not (only) how much you give ..., CESifo Working Paper No. 2944.

Erat, S., Gneezy, U., 2015. Incentives for creativity. Experimental Economics 18, 1–12.

Eriksson, T., Teyssier, S., Villeval, M.-C., 2009. Self-selection and the efficiency of tournaments. Economic Inquiry 47 (3), 530–548.

Eysenck, H. J., 1995. Genius: the natural history of creativity. Cambridge University Press, Cambridge, UK.

Fehr, E., Falk, A., 2002. Psychological foundations of incentives. European Economic Review 46 (4-5), 687–724.

Fehr, E., Gächter, S., 2000. Fairness and retaliation: The economics of reciprocity. Journal of Economic Perspectives 14 (3), 159–181.

Fehr, E., Gächter, S., 2002. Do incentive contracts undermine voluntary cooperation? IEW Working Paper No. 34.

Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: Experimental evidence. Econometrica: Journal of the Econometric Society 65 (4), 833–860.

Fehr, E., Schmidt, K., 1999. A theory of fairness, competition, and cooperation. The Quarterly Journal of Economics 114 (3), 817–868.

Fischbacher, U., 1999. z-Tree - experimenter's manual. Working Paper No. 21.

Florida, R., 2002. The Rise of the Creative Class. Basic Books New York.

Gill, D., Prowse, V., 2012. A structural analysis of disappointment aversion in a real effort competition. The American Economic Review 102 (1), 469–503.

Gneezy, U., List, J., 2006. Putting behavioral economics to work: Field evidence on gift exchange. Econometrica 74 (5), 1365–1384.

Greiner, B., 2004. An online recruitment system for economic experiments. In: Forschung und wissenschaftliches Rechnen 2003. Vol. 63. GWDG Bericht, pp. 79–93.

Guilford, J., 1967. The nature of human intelligence. New York: McGraw-Hill.

Guilford, J. P., 1959. Personality. McGraw-Hill, New York.

Harbring, C., Irlenbusch, B., 2003. An experimental study on tournament design. Labour Economics 10 (4), 443–464.

Harbring, C., Irlenbusch, B., 2011. Sabotage in tournaments: Evidence from a laboratory experiment. Management Science 57 (4), 611–627.

Hennig-Schmitt, H., Rockenbach, B., Sadrieh, A., 2010. In search of worker's real effort reciprocity – A field and a laboratory experiment. Journal of the European Economic Association 8 (4), 817–837.

Joussemet, M., Koestner, R., 1999. Effect of expected rewards on children's creativity. Creativity Research Journal 12 (4), 231–239.

Kessler, J. B., 2013. When will there be gift exchange? addressing the lab-field debate with a laboratory gift exchange experiment. mimeo.

Kim, K., 2006. Can we trust creativity tests? a review of the torrance tests of creative thinking (TTCT). Creativity Research Journal 18 (1), 3–14.

Kremer, M., Williams, H., 2010. Incentivizing innovation: Adding to the tool kit. In: Innovation Policy and the Economy, Volume 10. University of Chicago Press, pp. 1–17.

Kube, S., MarÃl'chal, M. A., Puppe, C., 2013. Do wage cuts damage work morale? Evidence from a natural field experiment. Journal of the European Economic Association 11(4), 853–870.

Kube, S., Marechal, M. A., Puppe, C., 2012. The currency of reciprocity: Gift exchange in the workplace. American Economic Review 102 (4), 1644–62.

Laske, K., Schroeder, M., 2015. Quantity, quality, and novelty: Direct and indirect effects of incentives on creativity. mimeo.

Lazear, E. P., 2000. Performance pay and productivity. American Economic Review 90 (5), 1346–1361.

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.

Runco, M., 1991. Divergent thinking. Ablex Publishing, Norwood, N J.

Shalley, C., Zhou, J., Oldham, G., 2004. The effects of personal and contextual characteristics on creativity: where should we go from here? Journal of Management 30 (6), 933–958.

Shalley, C. E., Oldham, G. R., 1997. Competition and creative performance: Effects of competitor presence and visibility. Creativity Research Journal 10 (4), 337–345.

Simonton, D. K., 2004. Creativity in Science: Chance, logic, genius, and zeitgeist. Cambridge University Press.

Torrance, E., 1968. Torrance tests of creative thinking. Personnel Press, Incorporated.

Torrance, E. P., 1998. Torrance Tests of Creative Thinking: Norms-technical manual: Figural (streamlined) forms A & B. Scholastic Testing Service.

Weitzman, M. L., 1998. Recombinant growth. The Quarterly Journal of Economics 113 (2), 331–360.

Woodman, R. W., Sawyer, J. E., Griffin, R. W., 1993. Toward a theory of organizational creativity. The Academy of Management Review 18 (2), 293–321.

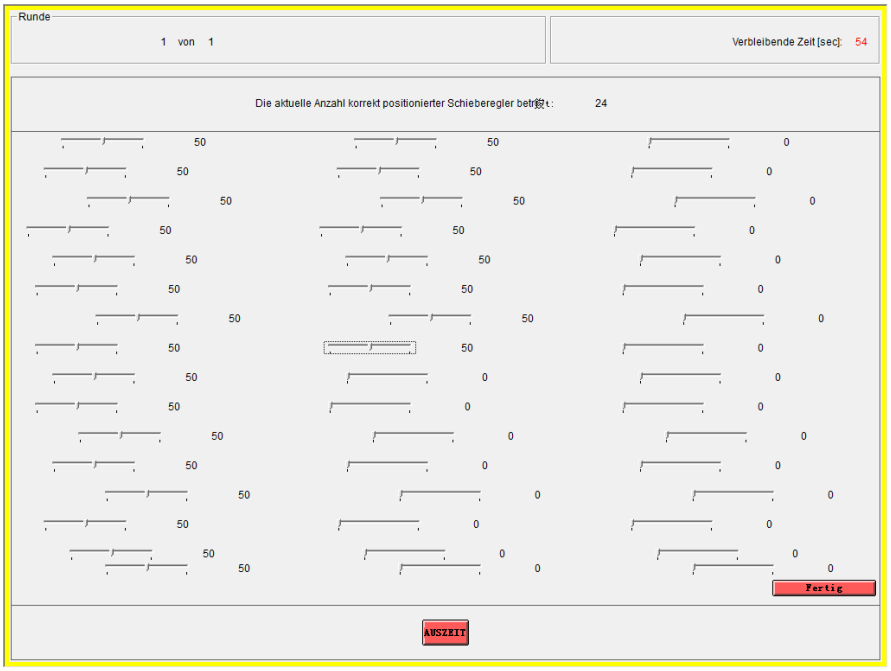
Figures and Tables

Figure 1: Overview of Payoffs by Treatment, Role, and Period

Work round	Role	Control	Gift for all (unconditional)	Tournament bonus for best 50% (conditional)
1	Employer Employee	300 300	300 300	300 300
Inter-vention		-	Bonus of 300 Thaler for all 4 employees	Bonus of 600 Thaler for best 2 (of 4) employees
2	Employer Employee	100 600	300 -200 = 100 300 + 300 = 600	300 -200 = 100 300 + 300 (EV) = 600
3	Employer Employee	300 300	300 300	300 300

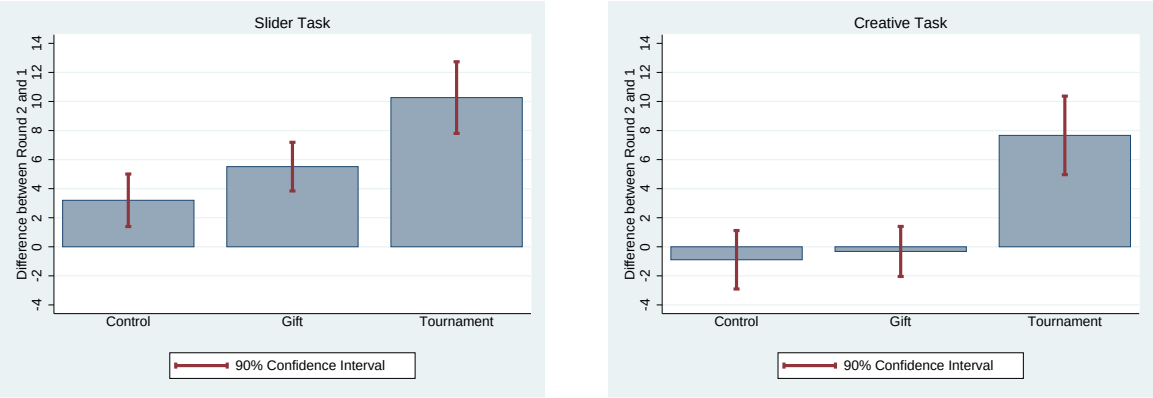
Note: 100 Taler translate to 1 Euro.
In the simple task, principals receive 5 Taler for each correctly positioned slider. In the creative task, principals receive 5 Taler for each per valid answer, 5 Taler for each new category, and 5 (10) Taler for each (very) original answer.
Agents earn an additional 5 Taler each time that they press the time-out button.

Figure 2: Screenshot of the Slider Task



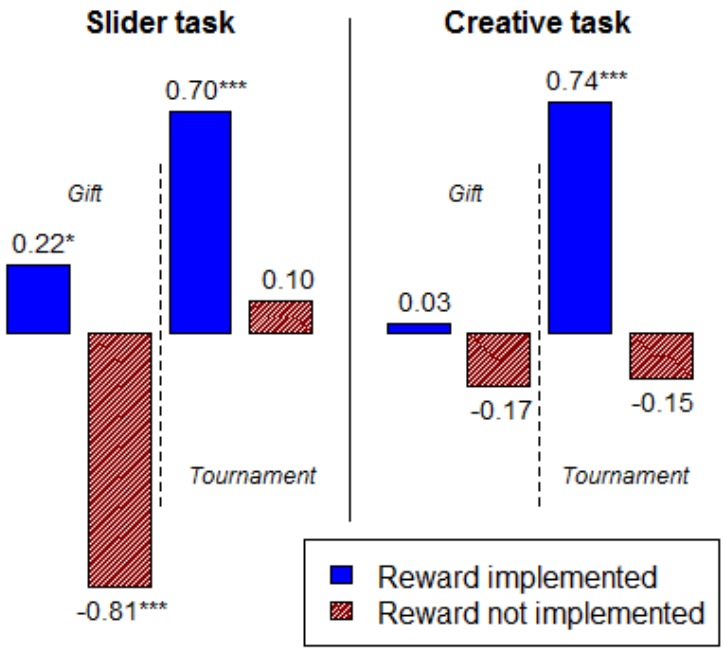
Note: The figure presents a screenshot of the slider task. The screen displays the remaining time and the number of correctly positioned sliders. The time-out button is displayed at the bottom of the screen.

Figure 3: Difference in Performance between Periods 2 and 1 by Treatment and Task



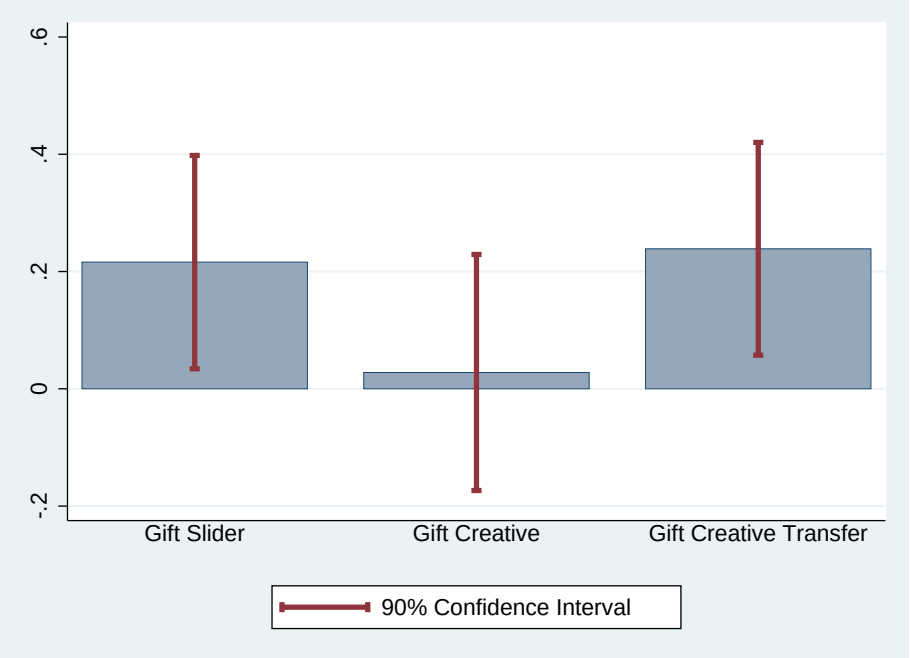
Note: The bars show the difference in mean (unstandardized) performance between Period 2 and Period 1. The whiskers depict 90% confidence intervals of paired t-tests. Note that the Y-axes are not comparable across tasks.

Figure 4: Effect Sizes for Positive and Negative Reward Decisions by the Principal



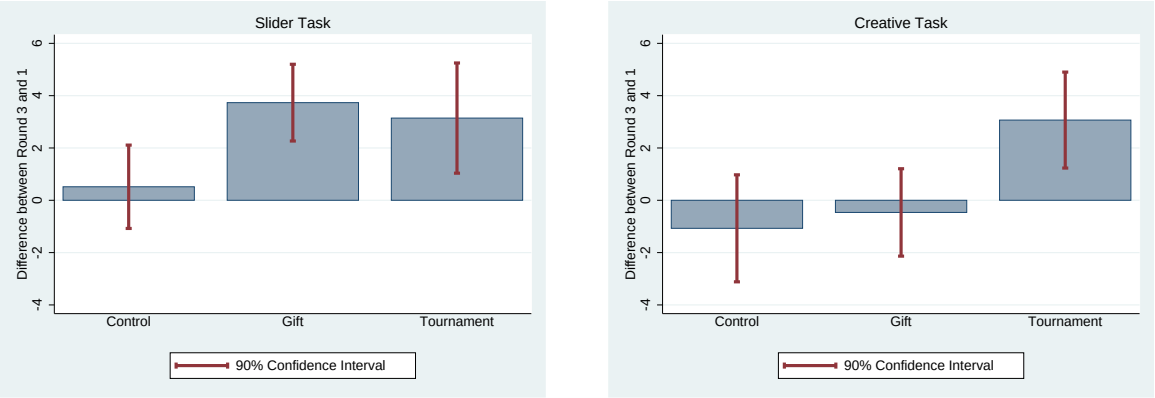
Note: The dependent variable is standardized performance of agents in Period 2. The bars show the estimated regression coefficients of separate OLS regressions. Performance is measured as the number of correctly positioned sliders in the simple task and as the score achieved in the creative task. The regressions control for baseline performance in Period 1. The respective control group (slider or creative) serves as the reference category. Statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5: Overview of Effect Sizes for All Gift Treatments



Note: The bars show the estimated regression coefficients of separate OLS regressions. Standardized performance (transfer) in Period 2 is the dependent variable. Performance is measured as the number of correctly positioned sliders in the simple task, as the score achieved in the creative task, and as the amount tranferred in the creative transfer treatments. The regression controls for baseline performance in Period 1. The respective control group serves as the reference category.

Figure 6: Difference in Performance between Periods 3 and 1 by Treatment and Task



Note: The bars show the difference in mean performance between Period 3 and Period 1. The whiskers depict 90% confidence intervals of paired t-tests. Note, this figure is on a bigger scale than Figure 3 to improve readability.

Table 1: Examples of Answers and Categories for the Unusual Uses Task by Item

	Paper	**Tin can**	**Cord**
Frequent Answers	Paper airplane Paper hat Toilet paper	Pen container Tin can phone Ball	Shoestrings Dog leash Fishing line
Frequent Categories	Toys Clothing Hygiene/Cleaning	Non-food container Communication Sport devices	Clothing accessories Leashes Fishing
Original Answers	Lampshade Filter Game of cards	Bedframe Animal house Insect trap	Pulley Rope bridge Bowstring
Very original Answers	Sound amplifier Pin wheel Artificial snow (for decoration)	Scarecrow Shower head Treasure chest	Trap (to play a trick) Straightening of acreages To cut a cake
Invalid Answers	Pencil Television Surfboard	Computer Window Shoes	Glasses Electric conductor Rope for bungee jumping

Table 2: Balance Table

	Main Treatments						Supplementary Treatments	
	Slider Task			Creative Task			Creative Transfer	
	C	T	G	C	T	G	C	G
# of Subjects	75	95	100	70	105	135	89	94
# Agents	60	76	80	56	84	108	71	76
# Rewarded Agents	60	56	60	56	60	56	71	68
Mean Age	24.9	23.1*	23.3	24.0	22.9*	22.8*	22.1	23.4
Share of Women	58%	52%	43%	52%	53%	57%	56%	59%
Economics major	50%	55%	57%	59%	68%	61%	66%	53%
Location								
Frankfurt	67%	64%	60%	71%	73%	71%	56%	54%
Mannheim	33%	36%	40%	29%	27%	29%	44%	46%
Baseline Performance	16.6	22.0**	18.9	16.8	17.8	16.0	19.8	20.3
	(12.5)	(12.8)	(14.1)	(11.7)	(10.9)	(10.7)	(11.7)	(10.0)

Note: Abbreviations: C = Control Treatment; T = Tournament Treatment; G = Gift Treatment.
 # Rewarded Agents reports the number of agents with a positive reward decision by the principal in the reward treatment groups. In the *Control* treatment it refers to the total number of agents. Hence, *# Rewarded Agents* presents the number of subjects per treatment group that we use in the statistical analysis. All stated means and percentages refer to the number of rewarded agents.
Baseline performance reports the mean (standard deviations in parentheses) of the number of correctly positioned sliders or of the Unusual Uses score in working period 1 – prior to any treatment intervention. Stars indicate the results from a unpaired t-test on the difference between performance in the respective treatment and control group. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Treatment Effects in Period 2

	I	II	III
Tournament Treatment	0.795***	0.736***	0.751***
	(0.173)	(0.143)	(0.143)
Tournament x Slider-Task	0.166	-0.058	-0.060
	(0.178)	(0.166)	(0.168)
Gift Treatment	-0.019	0.026	-0.006
	(0.143)	(0.099)	(0.103)
Gift x Slider-Task	0.375**	0.209*	0.206*
	(0.170)	(0.109)	(0.112)
Baseline		0.657***	0.638***
		(0.076)	(0.076)
Baseline x Slider-Task		0.002	0.028
		(0.094)	(0.094)
Intercept	-0.000	-0.000	0.874
	(0.093)	(0.062)	(0.595)
Controls	NO	NO	YES
Observations	348	348	348
R^2	0.146	0.546	0.564

Note: This table reports the estimated OLS coefficients from Equation 1. The dependent variable is the standardized performance in Period 2 and refers to the number of sliders moved in the simple task and the creativity score in the creative task. The coefficient on the *Tournament* (*Gift*) treatment represents the impact of the *Tournament* (*Gift*) on the creative task. The interaction effect between the respective treatment and the slider task documents whether the response to the *Tournament* (*Gift*) differs between the creative and the simple task. The impact of the tournament (*Gift*) on the simple task is equal to the sum of the coefficient on the treatment and its interaction effect.

The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Additional control variables are age, age squared, sex, location, field of study as well as a set of time fixed effects (semester period, semester break, exam period). Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effects Creative Transfer

	Amount Transferred	
Gift	0.224**	0.264**
	(0.103)	(0.126)
Baseline Transfer	0.653***	0.653***
	(0.063)	(0.061)
Intercept	-0.000	-0.677
	(0.077)	(0.851)
Controls	No	YES
Observations	139	139
R^2	0.515	0.540

Note: This table reports the estimated OLS coefficients from Equation 1. The dependent variable is the standardized amount transferred in Period 2. Amount transferred refers to the treatment in which agents work on the creative task, are informed about the number of points that they generated with their answers and can then decide how many of those points to transfer to their principal. The coefficient on the *Tournament* (*Gift*) treatment represents the impact of the *Tournament* (*Gift*) on the creative task. The interaction effect between the respective treatment and the slider task documents whether the response to the *Tournament* (*Gift*) differs between the creative and the simple task. The impact of the tournament (*Gift*) on the simple task is equal to the sum of the coefficient on the treatment and its interaction effect.

The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Additional control variables are age, age squared, sex, location, field of study as well as a set of time fixed effects (semester period, semester break, exam period). Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Dimensions of Creativity: Treatment Effects on Standardized Performance in Period 2

	Creative Task							
	Score	*Validity*	*Flexibility*	*Originality*	*Flexibility Rate*	*Originality Rate*	*Top Answers*	*Invalid Uses*
	I	II	III	IV	V	VI	VII	VIII
Tournament	0.753***	0.813***	0.607***	0.726***	-0.035	0.106*	0.199**	0.286
	(0.162)	(0.171)	(0.151)	(0.180)	(0.045)	(0.062)	(0.087)	(0.208)
Gift	-0.059	-0.027	-0.041	-0.143	0.014	-0.083	0.011	-0.178
	(0.124)	(0.131)	(0.120)	(0.150)	(0.048)	(0.068)	(0.087)	(0.187)
Period 1	0.614***	0.581***	0.598***	0.552***	0.058	0.081	0.197***	0.262***
	(0.077)	(0.078)	(0.071)	(0.087)	(0.063)	(0.089)	(0.061)	(0.065)
Intercept	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	172	172	172	172	152	152	172	172
R^2	0.500	0.474	0.526	0.389	0.068	0.117	0.179	0.190

Note: This table reports OLS estimates of Equation 1, where we regress (standardized) performance in Period 2 on baseline performance and treatment dummies. Column I reports results on the aggregated creativity score. Columns II, III and V report the results on the different standardized sub-dimensions of the creativity score. Column IV and VI displays treatment effects on the (unstandardized) flexibility and originality rate. The flexibility (originality) rate equals achieved flexibility (originality) points divided by the total number of subject's valid answers (subjects with zero valid answers were dropped in these columns). Column VII reports results for an assesment of the best creative answers from an evaluator blind to the treatments. Column VIII report results for invalid uses.
The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Additional control variables are age, age squared, sex, location, field of study as well as a set of time fixed effects (semester period, semester break, exam period). Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Translated Instructions (Original in German)

General Instructions

Please read the instructions carefully. If you have any questions, please raise your hand. Keep in mind that communication among participants is prohibited during the experiment. Please turn off your mobile phone and other electronic devices for the entire duration of the experiment. During the experiment, you will have the opportunity to earn money in the form of Taler. How many Taler you will earn depends on a random draw as well as on your decisions and the decisions of other participants. All Taler that you earn in the experiment will be converted to Euros at the end of the experiment. The exchange rate is

100 Taler = 1 Euro

At the end of the experiment you will receive the amount of money that you have earned during the experiment in cash. Your earnings will be rounded up to full 10-cent amounts. We would like to point out that your name is only required for the settlement of payments at the end of the experiment. Your name will not be connected to any decisions you make during the experiment. Your actions are completely anonymous.

Assignment of roles

At the beginning of the experiment a computer will randomly assign you the role of an "employer" or an "employee". You will keep this role for the entire experiment. Further, you will be randomly assigned to groups of 5 participants. One employer and four employees form one group. The groups will remain the same for the entire experiment.

Structure of the experiment

The experiment consists of two parts. Part 1 is the actual experiment. Employees only interact with their employer in Part 1. Part 2 consists of a series of further decision tasks. Only employees will be active in Part 2. The instructions for Part 2 of the experiment will be shown on your computer screen once Part 1 is finished. After to Parts 1 and 2, both employees and employers will be asked to fill out a questionnaire. Please find the instructions for Part 1 below.

Part 1 Instructions

Fixed payment

Part 1 of the experiment will be carried out in three rounds. In the first round employers and employees will receive a fixed payment of 300 Taler. The amount of the fixed payment for the second and third round will be displayed on the computer screen shortly before the respective round starts.

Variable payment

Employers earn additional money. How much they earn is determined by the work performance of the four employees in her group.[34] Towards that end, employees are asked to work on a simple task in each round. The task will be the same in every round. Employers' additional earnings depend on the overall performance of her four employees in all three rounds. The employer will not receive any information on agents' performance or her own earnings until the very end of the experiment. Employees are free to decide how much effort they want to exert and thereby how much money they want to earn for their employer. Employees' own earnings will not be influenced by their work performance. Employees do, however, have the option to press a time-out button while they work on the task. This button will lock the screen for 20 seconds so that they cannot work on the task during that time period. Employees receive 5 Taler for each time that they press the time-out button. The button can no longer be pressed when there are less than 20 seconds left in a round. Employers are also allowed to work on the task in all three rounds. Their performance, however, will neither influence their own payment nor their employees' payments. The employer can also press the time-out button. Pressing the button, however, does not translate into extra income for the employer. The employer cannot earn additional Taler during time-out.

[The original instructions had the description of the creative or the simple, routine task at this point.]

The employer will not receive any information related to how many points each of her agents generated. Her total payoff will only be revealed at the end of the experiment. Employees will also not be informed about the total amount of Taler generated by themselves or other group members. All participants will be called individually for payment. Please remain in your seat at the end of the experiment until your seat number has been called.

[Further instructions, depending on the treatment were displayed on the computer screen.]

[34] For simplicity we use the feminine (she, her) to describe employers, but this does, of course, refer to both sexes and employers will be both male and female.

Task description: Creative task

As described above, the first part of the experiment consists of three rounds. In each round, participants will work on a task that requires creativity.

Example of the task

Please list as many, as different and as unusual uses for a rubber tire as you can think of. Do not restrict yourself to a specific size of a tire. You can also list uses that require several tires. Do not restrict yourself to uses you are familiar with, but think of as many new uses as possible!

You will receive the same task in each round with varying objects. You will have three minutes per round. There will be a break of a few minutes after every round.

Evaluation of the task

The employer receives 5 Taler per valid answer. Answers are "valid" when they are practicable and when their realization is at least vaguely conceivable. Please describe the possible use in a few words if necessary (Using the example of the rubber tire: "sled" or "flower box" are clear answers, whereas "target" would require further explanation such as "ball game with tire as target".).

For original (rare) answers, the employer receives 5 extra Taler. For very original (very rare) answers, the employer receives 10 extra Taler. An answer is considered (very) "original" if only (very) few people think of it. To this end, the answers are compared to a catalog of answers that is based on the answers of more than 100 test persons.

Furthermore, answers will be assigned to different answer categories. The employer receives 5 extra Taler for each category answers fall into. Using the example of the rubber tire: "car tire" and "bicycle tire" belong to the category "tires as wheels" and result in 5 extra Taler. The answer "swing seat" is a different category (category "toys") and yields an additional 5 Taler.
At the end of the experiment, we will ask you to fill out a brief questionnaire while we evaluate the answers of all employees. We will calculate the employer's variable payment from the total score of all four employees over the three rounds.

Task description: Simple, routine task

In every round you will work on a task that requires you to use your mouse to move "sliders" on the computer screen. You will see 48 sliders on your screen in each round. These sliders have to be moved to position 50. Each slider starts at position 0 and can be moved anywhere up to position 100. The number displayed to the right of each slider indicates the slider's current position. You can adjust the slider as often as you like. Your working time is set to 3 minutes (180 seconds) per round. If you have moved all 48 sliders to the correct position before the end of a round, you will automatically be directed to a new screen with 48 additional sliders. Thus, you can move a maximum of 96 sliders to the correct position during the 180 seconds of each round. During each round, you will see some information at the top of your screen. You will see the round number, the remaining time for the current round, and the number of sliders that you have moved to position 50 in the current round so far. Before the first round starts, you will be able to test the task for 60 seconds. Your performance in the test round will not affect your own or your employer's payoff. The employer will receive 5 Taler for each slider that you correctly position in the three rounds. At the end of the experiment, the number of correctly positioned sliders of all four employees of a group across all three rounds are summed up to calculate the variable payment of the employer.

Webappendix

Table 6: Overview of Payoffs by Treatment, Role, and Period

Payments in Taler						
	Control		Gift		Tournament	
	Principal	Agent	Principal	Agent	Principal	Agent
Period 1 - Fixed Wage	300	300	300	300	300	300
Period 2						
Fixed Wage	100	600	300	300	300	300
Reward Costs(-)/Benefits(+)	-	-	-200	+300	-200	+0/600[a]
Total (Expected) Payoff	100	600	100	600	100	600[b]
Period 3 - Fixed Wage	300	300	300	300	300	300

Note: [a] Tournament winners (best 50%) receive a bonus of 600 Taler and tournament losers (worst 50%) receive nothing.
[b] Assuming risk neutrality and a 50% chance of winning, a subject's expected earnings in the tournament treatment at the beginning of the second period as well as average earnings at the end of the second period are 600 Taler.
The experimental currency unit "Taler" was converted into Euros at an exchange rate of 100 Taler = 1 Euro.

Variable Payments (in Taler):	
Principal:	
Per Slider	5
Per Valid Answer	5
Per Category	5
Per Original Answer	5
Per Very Original Answer	10
Agent:	
Per Time-out	5

Table 7: Balance Table

	Main Treatments						Supplementary Treatments						
	Slider Task			Creative Task			Slider Task	Creative Task	Creative Transfer		Mixed Task (SCS)	Mixed Task (CSC)	
	C	G	T	C	G	T	F	F	C	G	T	T	**Total**
# of Subjects	75	100	95	70	135	105	80	90	89	94	94	96	1123
# Agents	60	80	76	56	108	84	64	72	71	76	75	77	899
# Rewarded Agents	60	60	56	56	56	60	56	68	71	68	55	46	712
													Mean
Mean Age	24.9	23.3	23.1*	24.0	22.8*	22.9*	23.2	23.4	22.1	23.4	22.4**	21.6***	23.1
Share of Women	58%	43%	52%	52%	57%	53%	57%	38%	56%	59%	58%	50%	53%
Economics major	50%	57%	55%	59%	61%	68%	64%	49%	66%	53%	49%	67%*	58%
Location													
Frankfurt	67%	60%	64%	71%	71%	73%	64%	59%	56%	54%	27%***	26%***	58%
Mannheim	33%	40%	36%	29%	29%	27%	36%	41%	44%	46%	22%	22%	34%
Heidelberg	0	0	0	0	0	0	0	0	0	0	51%***	52%***	7%
Baseline Performance	16.6	18.9	22.0**	16.8	16.0	17.9	20.3	17.9	19.8	20.3	19.5	17.8	18.6
	(12.5)	(14.1)	(12.8)	(11.7)	(10.7)	(10.9)	(11.4)	(10.3)	(11.7)	(10.0)	(13.1)	(10.4)	(11.6)

Note: Abbreviations: C = Control Treatment; G = Gift Treatment; T = Tournament Treatment; F = Feedback Treatment; SCS = Slider-Creative-Slider Treatment; CSC = Creative-Slider-Creative Treatment.

 # Rewarded Agents reports the number of agents with a positive reward decision by the principal in the reward treatment groups. In the *Control* treatment it refers to the total number of agents. Hence, *# Rewarded Agents* presents the number of subjects per treatment group that we use in the statistical analysis. All stated means and percentages refer to the number of rewarded agents.

Baseline performance reports the mean (standard deviations in parentheses) of the number of correctly positioned sliders or of the Unusual Uses score in working period 1 – prior to any treatment intervention. Stars indicate the results from a unpaired t-test on the difference between performance in the respective treatment and control group. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Sensitivity Analyses

	Score	Validity	Flexiblity	Originality 10%	Originality 5%	Originality 1%	Invalid Uses
Gift	0.026	0.054	0.055	-0.101	-0.075	0.237	-0.112
	(0.121)	(0.125)	(0.116)	(0.141)	(0.143)	(0.190)	(0.166)
Tournament	0.736***	0.789***	0.589***	0.683***	0.672***	0.747**	0.294
	(0.160)	(0.168)	(0.146)	(0.190)	(0.193)	(0.194)	(0.198)
Baseline Performance	0.657	0.628***	0.628***	0.512***	0.564***	0.557**	0.255***
	(0.076)	(0.099)	(0.092)	(0.083)	(0.090)	(0.117)	(0.080)
Intercept	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.473	0.433	0.5	0.309	0.311	0.257	0.128
Observations	172	172	172	171	171	171	172

Note: This table reports OLS estimates of Equation 1, where we regress standardized performance in Period 2 on baseline performance and treatment dummies. Column I reports results on invalid uses. Columns II to VI report the results on the different sub-dimensions of the creativity score. Column IV to VI report robustness for the calculation of the originality score (given exactly point for each *original* answer given by less than 10%, 5% or 1% of agents).
The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Treatment Effects in Period 3

	Slider		**Creative**	
	I	II	III	IV
Gift	0.282**	0.292**	0.006	-0.005
	(0.112)	(0.112)	(0.116)	(0.116)
Tournament	0.228		0.364***	
	(0.154)		(0.126)	
Creative Transfer			-0.039	-0.036
			(0.122)	(0.122)
Creative Transfer x Gift			0.400***	0.397***
			(0.120)	(0.120)
Tournament Winner		0.497***		0.488***
		(0.188)		(0.159)
Tournament Loser		-0.019		0.243*
		(0.162)		(0.159)
Controls	YES	YES	YES	YES
Baseline	YES	YES	YES	YES
Intercept	YES	YES	YES	YES
N	176	176	311	311
R^2	0.734	0.749	0.543	0.547

Note: This table reports OLS estimates of standardized performances in Period 3 analogous to Equation 1. Performance is measured as the number of correctly positioned sliders, as the score achieved in the creativity task, and as the amount transferred in the Creative Transfer treatments.
The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Additional control variables are age, age squared, sex, location, field of study as well as a set of time fixed effects (semester period, semester break, exam period). Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Period 2 and Period 3 Effects of the Feedback Treatment

	Slider-Task	**Creative-Task**
Feedback Period 2	0.184*	0.182*
	(0.108)	(0.102)
Feedback Period 3	0.118	0.171
	(0.132)	(0.129)
Positive Relative Feedback Period 3	0.303	0.319**
	(0.196)	(0.148)
Negative Relative Feedback Period 3	-0.028	0.035
	(0.135)	(0.144)
Controls	YES	YES

Note: This table reports the estimated OLS coefficient estimates for the *Feedback* Treatment for Periods 2 and 3. Period 3 effects are split up by the whether person learned that they did or did not belong to the 50% top performers. Feedback effects for Period 2 are regression coefficients from specification IV in Table 3 while post-treatment effects for Period 3 are from Table 9. The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Control variables include observations from the *Gift* treatment, the *Tournament* treatment, and the *Control* group. Additionally, we control for age, age squared, sex, location, field of study as well as a set of time fixed effects (semester period, semester break, exam period). The *Control* group serves as the reference category. Heteroscedasticity-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Post-treatment Effects of the Tournament in Period 3 by Task

	Slider Task		**Creative Task**		**Mixed Tasks** **Slider-Creative-Slider (SCS)**		**Creative-Slider-Creative (CSC)**	
	I	II	III	IV	V	VI	VII	VIII
Tournament	0.245*		0.340***		0.111		0.055	
	(0.137)		(0.126)		(0.109)		(0.121)	
Tournament Winner		0.534***		0.449***		0.064		0.091
		(0.180)		(0.170)		(0.122)		(0.146)
Tournament Loser		-0.030		0.236*		0.158		0.007
		(0.146)		(0.134)		(0.142)		(0.132)
Standardized Performance in Period 1	0.888***	0.848***	0.693***	0.670***	0.888***	0.891***	0.648***	0.648***
	(0.060)	(0.063)	(0.080)	(0.086)	(0.052)	(0.053)	(0.089)	(0.090)
Intercept	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.073)	(0.073)	(0.093)	(0.094)	(0.073)	(0.073)	(0.094)	(0.094)
Observations	116	116	116	116	115	115	102	102
R-squared	0.679	0.708	0.529	0.535	0.734	0.735	0.503	0.504

Note: This table reports OLS estimates of standardized performances in Period 3. Performance is measured as the number of correctly positioned sliders in the simple task and as the score in the creative task, respectively.
The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament as well as agents from the supplementary treatments SCS and CSC. Agents with negative reward decisions are not part of this analysis. Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Treatment Effects in Period 2 – Controlling for Number of Breaks Taken

	Slider Task I	Creative Task II
Gift Treatment	0.019	-0.056
	(0.056)	(0.110)
Tournament Treatment	0.156**	0.269**
	(0.064)	(0.125)
1 to 3 Breaks (Period 1)	-0.076*	-0.188
	(0.117)	(0.169)
4 to 7 Breaks (Period 1)	-0.362*	-0.122
	(0.186)	(0.168)
8 or more Breaks (Period 1)	-0.630**	-0.573***
	(0.250)	(0.196)
Difference between the	-0.734***	-0.204***
Breaks in Periods 2 and 1	(0.146)	(0.061)
1 to 3 Breaks (Period 1)	0.530***	-0.031
× Break Difference	(0.149)	(0.071)
4 to 7 Breaks (Period 1)	0.461***	0.030
× Break Difference	(0.147)	(0.078)
8 or more Breaks (Period 1)	0.414***	-0.082
× Break Difference	(0.146)	(0.080)
Standardized	0.767***	0.721***
Performance in Period 1	(0.095)	(0.077)
Intercept	0.141	0.275*
	(0.132)	(0.163)
Observations	176	172
R^2	0.915	0.677

Note: In order to create a more realistic work environment in which agents face opportunity costs of working, we implemented a time-out button (see Section 3.2) which was displayed at the bottom of the screen during all working periods. Each time agents clicked the time-out button, the computer screen was locked for 20 seconds, prohibiting the entry of creative ideas or the movement of sliders, and 5 Taler were added to the agent's payoff.

This table reports OLS coefficient estimates from our main model specification (Equation 1, with the number of breaks in period 1, the change in breaks between periods 1 and 2, as well as interactions between the two as additional controls. Standardized performance in period 2 is the dependent variable. Performance is measured as the number of correctly positioned sliders in the simple task and as the score in the unusal uses, creative task, respectively.

The estimation includes all agents from the Control Group as well as agents from treatment groups where the principal decided to institute the tournament/gift. Agents with negative reward decisions are not part of this analysis. Heteroscedasticity-robust standard errors including a degree of freedom correction are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.