

TI 2016-035/VII
Tinbergen Institute Discussion Paper



Incentivizing Creativity: A Large-Scale Experiment with Tournaments and Gifts

Christiane Bradler¹

Susanne Neckermann^{2,1}

Arne Warnke¹

¹ ZEW Centre for European Economic Research, Mannheim, Germany;

² Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, the Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

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

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

April 2016

Abstract

This paper reports the results from a large-scale laboratory experiment investigating the impact of tournament incentives and wage gifts on creativity. We find that tournaments substantially increase creative output, with no evidence for crowding out of intrinsic motivation. By comparison, wage gifts are ineffective. Additional treatments show that it is the uncertain mapping between effort and output that inhibits reciprocity. This uncertainty is prevalent in creative and other complex tasks. 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. Specific 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".

[†]ZEW Centre for European Economic Research Mannheim. E-mail: bradler@zew.de.

[‡]Erasmus University Rotterdam, Tinbergen Institute, and ZEW. E-mail: neckermann@ese.eur.nl.

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

1 Introduction

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). This trend requires firms to adapt and change in many ways. One particular challenge relates to incentivizing workers to perform well in these types of jobs. To date little is known about how creative performance responds to incentives.¹ By comparison, there is a long tradition in economics that investigates the impact of work incentives on motivation and productivity in simple and routine tasks. The purpose of this paper is to gain a deeper understanding of the incentive response function of creativity by investigating the extent to which lessons learned from simple tasks can be generalized to creative tasks.

Dozens of studies have explored the impact of financial incentives on routine tasks, confirming what standard economic theory predicts: financial incentives have a positive effect on performance 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).² But even as early as 1999, Prendergast noted that these types of routine jobs are not very common.

Explicit financial incentives are, of course, not the only way to trigger workers' performance. An established literature on gift exchange suggests that workers reciprocate wage gifts with higher effort (Akerlof, 1982). This 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). Like the literature on financial incentives, these studies have also almost exclusively utilized simple, routine tasks.

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

¹The nascent literature on incentivizing creativity in economics will be discussed below.

²In the field, researchers observed work performance when, for instance, installing wind shields (Lazear, 2000), picking fruits (Bandiera et al., 2005), or planting trees (Shearer, 2004); in the laboratory, subjects have been rewarded for typing letters (Dickinson, 1999), cracking walnuts (Fahr and Irlenbusch, 2000), or filling envelopes (Falk and Ichino, 2006). A main advantage of these tasks is that they offer a precise and easily observable measure of the quantity (and the quality) of workers' output.

jobs which rely on creativity.³ To date, economic theory mostly abstracts from the nature of the task and predicts uniform effects of rewards across different tasks. One exception is the literature on motivational crowding out that suggests that explicit incentives can sometimes be counterproductive. This is especially likely for tasks which are cognitively complex or high in intrinsic motivation, both of which are attributes of creative tasks (Bonner et al., 2000; Camerer and Hogarth, 1999; Amabile, 1996; Shalley et al., 2004). The idea is that the provision of financial incentives can “crowd out” the intrinsic motivation to perform the task (Deci et al., 1999).⁴ So far, the economic literature on crowding out has primarily focused on intrinsically motivated activities that are not creativity-oriented (e.g., Frey and Jegen, 2001, or Gneezy et al., 2011 for overviews) and still debates the existence of crowding out effects (e.g., Fang and Gerhart, 2012; Fehr and Falk, 2002; Charness and Gneezy, 2009). Gifts might be effective in situations where crowding out effects can be expected to be strong, for instance for creative work, yet there is no previous literature exploring this hypothesis.

This paper reports the results from a laboratory experiment with more than 1000 subjects that addresses these open issues. Subjects were randomly assigned to groups of five, with one principal (“employer”) and four agents (“employees”). The four agents worked for three periods for their principal on either a creative or a routine task. Our basic design is 2 x 3: the routine and the creative task with three treatments (control, gift, tournament) each. We also ran a host of supplementary treatments that serve to elucidate the underlying mechanisms. In all treatments, the principal’s payoff depended on the output produced by their four agents, who received an exogenously assigned fixed wage in each period. After agents worked on the task for one period, principals could decide whether or not to provide additional rewards to their agents (at their own expense and without knowledge about agents’ performance in Period 1). In the *Tournament* treatment, the principal could provide an additional monetary prize to the 50% best performing agents in her group. In the *Gift* treatment, the principal could opt for a monetary gift, half as large as the tournament prize, to all four agents in her group. In the third period, subjects were asked to work for the principal for one more round without additional rewards.

The experiment was run as a three period design to provide us 1) with a baseline measure of performance (capturing agents’ ability, intrinsic motivation to perform the task, and

³According to the dominant scholarly definition, creativity is defined as the production of ideas, solutions, or products that are novel (i.e., original) and appropriate (i.e., useful) in a given situation (Amabile, 1997). In contrast, a task can be defined as routine “if it can be accomplished by machines following explicit programmed rules.” (Autor et al., 2003, p. 1283). The addition of “simple” refines this latter definition by restricting the set of tasks to those that are simple to understand and perform, i.e., that do not require much instruction, skill, or prior knowledge.

⁴According to self-perception theory (Bem, 1972), a reward causes a shift in people’s perception of why they perform a task: own behavior is attributed to the reward and not to the enjoyment of the activity itself. As a result, if a reward is subsequently removed, individuals are less motivated to work on the task (Deci, 1971; Lepper et al., 1973; Amabile, 1988, 1996; Joussemet and Koestner, 1999; Deci et al., 1999, among others). Other reasons for a negative effect of rewards on performance include the exertion of “too much effort” or an increase in self-consciousness beyond an optimal level (e.g., Camerer and Hogarth, 1999 or Ariely et al., 2009).

other-regarding preferences towards the principal), 2) with a measure of performance under both types of rewards schemes for both tasks,⁵ and 3) with a measure of whether or not treatments affected intrinsic motivation (which can only be captured in a subsequent period without rewards). Both reward schemes were designed to have identical costs to the principal and overall benefits to the group of workers; they differed only in the distribution of benefits among workers. This feature allows conclusions about which of the two schemes is preferable given a fixed budget for rewards.

In addition to the two monetary reward treatments, we ran a control group for each task which allows us to account for learning and fatigue and to standardize performance across the two tasks, rendering the effect sizes comparable.

We find a substantial and positive incentive effect of the *Tournament* in both tasks (routine and creative) in the second period. The effect sizes are of similar magnitude in the two tasks, suggesting that performance in both tasks is equally sensitive to competitive incentives. Subsequent to learning whether they were winners or losers at the end of Period 2, winners in the *Tournament* outperformed comparable others in the control group in Period 3 even though there was no subsequent tournament prize at stake. Losers, on the other hand, returned to their baseline level of performance. This is evidence against the notion that financial rewards crowd out intrinsic motivation and thus creative performance. To the contrary, it suggests that performance-dependent rewards in the form of tournament incentives increase creative performance and even have long-lasting effects on tournament winners.

Interestingly, 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, agents do not reciprocate the gift in the creative task. We discuss various potential explanations for this asymmetry and show that the absence of reciprocity in the creative task is driven by employees' lack of ability to finetune their back transfer to the principal. Creative tasks are characterized by this inability since the exact value of an idea is typically somewhat uncertain and only becomes apparent with time, say, after an idea is implemented. The same was true in our experiment where the value of an idea to the principal depended on the originality rating of an idea. Overall our results suggest that creative tasks respond well to performance incentives and that responses to rewards are very similar in creative and routine tasks; however, certain features inherent in creative work affect the effectiveness of gifts as rewards. Therefore, wage gifts for employees in creative tasks might not effectively boost creativity in practice.

⁵We 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.

This study contributes to the small but growing literature in economics that studies the impact of rewards on creativity. 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 find evidence for choking under pressure.⁶

We extend this nascent economic literature on creativity as well as the literature on incentive provision in five 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.⁷ 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 look into whether the incentive-response function differs between our creative and our simple, routine tasks. This comparison provides first insights into whether the lessons learned with simple tasks generalize to creative tasks, and, hence, whether an independent literature assessing incentives for creativity is needed. Fourth, we introduce the Unusual Uses task into the experimental literature on creativity. In our view, this task captures central elements of day-to-day creativity in organizations well (Woodman et al., 1993).⁸ 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

⁶Another 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.

⁷We used a tournament scheme 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).

⁸The task has also been used by Dutcher (2012) in his study on the effects of telecommuting on productivity.

and in corporate idea-suggestion systems. In that sense, the paper complements the existing literature that mostly focuses on blue-sky creativity or on tasks that involve very little creativity such as those that involve pattern recognition. An 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 performance. Finally, we provide another data point to the discussion on whether or not it is possible to foster creative performance through financial rewards.

The paper is structured as follows. Section 2 describes the experimental set-up, the tasks, and the treatments. Section 3 presents our main results, Section 4 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 The Experiment

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

2.1 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 1 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.⁹ 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

⁹Keyboards were disconnected during the task to prevent the usage of the arrow keys.

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 allotted 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).¹⁰ 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; 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 Unusual Uses task 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: (1) paper, (2) tin can, (3) cord.¹¹

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.¹² 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.¹³ In the experiment,

¹⁰In order to assess overall creative potential, the *Torrance Test of Creative Thinking* also includes a number of figural elements that require drawing skills. The more specific Unusual Uses task, however, best captures the type of creativity that we want to study.

¹¹Controlling for order effects is not important in our design as we use the same order in the control group and only look at changes in performance between periods and between treatment and control groups.

¹²The 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.

¹³Usual 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.

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.¹⁴ 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 carefully acquainted with the scoring procedures and blind to the treatments. In order to test whether the creative task is more intrinsically motivating than the routine task, we had 100 subjects work on both tasks and rate their interest in the tasks on Likert-scales from 1 to 7. According to this assessment, the creative task was rated as statistically significantly more interesting than the routine task (Wilcoxon rank-sum test, $p=0.02$).

2.2 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 considerations as it allows 1) for voluntary financial transfers from the principal to the agent, and

Excluding usual uses, however, does not alter any of the results reported below (results available upon request).

¹⁴The answers gathered during the experiment itself allowed us to update the rating scale to more accurately 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).

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. In each period, employees received a fixed wage that was exogenously set by the experimenter and announced at the start of the respective period. In the two reward treatments, principals could opt for or against a financial gift or a tournament before the start of Period 2 (details below). In the *Gift* treatment, Period 2 endowments were augmented by the amount of the gift when the principal opted for the gifts. In the *Tournament* treatment, agents could receive a tournament prize at the end of Period 2 if the principal opted for the tournament scheme and if their performance was above average. In all treatments, Period 3 was identical to Period 1 in that endowments were fixed and there were no additional rewards or tournaments. 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 and post-treatment performance under a fixed wage in Period 3.¹⁵ Period 3 is important as detrimental effects of rewards on intrinsic motivation cannot necessarily be detected when rewards are present as the increase in monetary incentives might outweigh the reduction in intrinsic motivation. After the three working periods, agents completed a couple of brief decision tasks; questions about their socio-demographic characteristics such as gender, field of study, level of education, high school grade and leisure activities; and 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.¹⁶ 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,

¹⁵As agents work for a fixed wage in Period 1, their effort does not affect their own payoff. It does, however, affect the payoff of their principal. Therefore, baseline performance in Period 1 provides us with a joint measure of intrinsic motivation for the task, their altruism towards the principal, and a subject's ability.

¹⁶ The exchange rate was 100 Taler = 1 Euro.

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).¹⁷

2.3 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.

In the *Control* group, agents were paid a fixed wage in each period and principals were not able to implement rewards. 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. The same is true for the *Control* group with the exception of Period 2 in which the fixed payment was doubled to mirror (expected) payoffs in the treatment groups (see below). 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.¹⁸ In the *Tournament* treatment, the principal could also transfer a total of 1200 Taler to her four agents at a cost of 200 Taler to herself. However, the payment structure was different.

¹⁷ 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.

¹⁸The use of efficiency factors is common practice in the experimental literature on gift exchange (see, for 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.

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.

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.

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 receive 300 Taler.¹⁹ 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).²⁰ This procedure ensures that any performance differences between the treatments and the control group are solely driven by the rewards and not by other factors such as distributional concerns or income effects.²¹ Table 2 in the Appendix provides an overview of the fixed and variable pay components for all periods, treatments, and roles.

2.4 Hypotheses

Before we present our results, we briefly summarize the theoretical predictions for our treatments.

Tournament treatment. According to standard economic theory, a tournament incentive should

¹⁹On top of their endowment, principals earn additional money from the performance of their agents.

²⁰If 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.

²¹See, 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.

increase average effort and performance (Alchian and Demsetz, 1972; Lazear and Rosen, 1981). Since economic theory tends to abstract from the type of task, this should be true in both the creative and the routine task. Theories on crowding-out, on the other hand, suggest that intrinsic motivation might be impaired by the reward (e.g., Deci, 1972; Lepper et al., 1973; Deci et al., 1999). This holds in particular for the creative task that is more intrinsically motivating. When crowding out is strong, the reduction in intrinsic motivation would outweigh the incentive effect and Period 2 performance would fall. Otherwise, there would be a positive response to the tournament incentive in Period 2, and a reduction in intrinsic motivation – if permanent – can be detected in Period 3, where the tournament reward is no longer at stake (Frey and Jegen, 2001; Bowles and Polania-Reyes, 2012).²²

Gift treatment. Established theories on reciprocity predict that agents reciprocate principals' gifts with increased effort (Akerlof, 1982; Levine, 1998; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). Thus, we expect a positive performance response to the gift in both tasks. Depending on how sustainable these responses are, this effect might carry over to Period 3.²³

2.5 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.

²²Unlike other models of crowding out, in Benabou and Tirole's (2006) model, crowding out only persists as long as the incentives are in place. Crowding out that is temporary and tied to the presence of incentives cannot be directly identified in our design. Most discussions of crowding out (e.g. in the education literature) focus on the permanent crowding out as a consequence of temporary incentives. It is that phenomenon that we attempt to capture in this paper.

²³Theoretically, agents could also reciprocate the investment in the tournament. We will address this point in the results section.

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.

3 Main Results

3.1 Descriptive Statistics and Methods

Table 3 provides an overview of the descriptive statistics split up by treatment and task. Although we don't discuss supplementary treatments until later, for completeness they are included in the balance table. Overall, our sample contains 1123 subjects: 224 employers and 899 employees.²⁴ 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.²⁵ 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. We also have data for 364 additional agents with positive reward decisions in supplementary treatments that we describe further below. The treatments are largely balanced with regard to the location of the experiment, gender, age, and field of study, albeit some differences 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 3 displays means and standard deviations of the baseline performance in Period 1 across treatments and tasks. Baseline performance provides a measure of agents' motivation to work on the task in the absence of rewards and comprises agents' ability, intrinsic motivation for the task, as well as their desire to benefit the principal. 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 2 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

²⁴We had to exclude 9 employees from the analysis due to an insufficient knowledge of the German language.

²⁵Principals did not receive any information on agents' performance in period 1 and agents knew this. Thus, as expected, our data does not reveal a significant relationship between reward implementation and baseline performance. We will report statistics on principals' decisions regarding reward implementation as well as agents' behavior in case of reward denial further below.

baseline performance in the analyses that follow and use the change in performance between Periods 1 and 2 as the outcome variable.

Figure 2 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 these two rounds. 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 probably a consequence of either learning or a response to the higher fixed wage in period 2.²⁶ 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. Below, we will corroborate these findings with more detailed analyses.

3.2 Further Analyses

In the analyses that follow, we compare the change in performance of individuals in the treatment groups with that of individuals in the respective control group. We make performance comparable between the different periods (and thus different items on the creative task) and the two different tasks by standardizing performance.²⁷ Further, we address potential mean reversion by controlling for baseline performance. This also allows us to exploit the panel structure of our experiment.

In order to assess the effect of the different rewards on performance in period 2, we fit the following regression model using ordinary least squares (OLS):

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

²⁶The higher fixed wage mirrors the wage of agents in groups with a positive reward decision by the principal (at least in expectation). This allows us to disentangle reward effects from endowment effects. This is of particular interest for the *Gift* treatment as it disentangles reciprocity from the response to the wage increase per se (Charness, 2004).

²⁷We 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.

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.²⁸ Column I presents the most parsimonious specification and does not control for baseline performance. Column II of Table 4 shows the results for equation 1. Column III adds additional control variables including age, age squared, sex, field of study, time period (semester, exam period, semester break), and location. Adding these controls does not alter the results.²⁹ Similarly, the results are robust to the exclusion of baseline performance as a control (see Column I of Table 4). When discussing the results below we refer to the specification in Column III which includes all control variables. Column IV reports the results of a supplementary treatment that will be discussed below (Section 4.1).

As was predicted by standard agency theory, the *Tournament* treatment has a large and statistically significant positive effect on performance in both tasks. This effect is substantially larger than that of the *Gift* treatment (Wald test, $p < 0.01$ for both tasks). 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: Agents increase their performance by approximately 0.7 standard deviations as compared to the control group in both the creative and the simple task. This is first evidence against crowding out as crowding out would have implied a smaller performance increase (if not a performance decrease) in the intrinsically motivating creative task. The analysis of performance in Period 3 below will further corroborate this finding.³⁰

While the *Tournament* worked well in stimulating creative performance, the *Gift* did not. The simple task shows that this absence of reciprocity is not driven by the gift being too small or too unimportant to have any effect. In the simple task the *Gift* induced an economically and statistically significant effect ($p < 0.05$). The effect size of 0.2 standard deviations is small relative to that of the tournament, but well in line with what other studies on reciprocity have

²⁸The results are robust to using cluster-robust standard errors (by session) that control for potential intra-session correlation.

²⁹The regressions in Columns III and IV also include observations from a supplementary feedback treatment, discussed below, that was run concurrently with the main treatments. We include observations from this treatment in order to estimate the coefficients of the other control variables, introduced in column III, with greater precision. 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.

³⁰Also, responses to the tournament do not differ between agents with above- and below-average baseline performance. Results available upon request.

found.³¹ The absence of reciprocity in the creative task is surprising and suggests that the responsiveness to rewards differs between simple and creative tasks, at least with respect to gifts. Below we will look further into this asymmetry in an effort to uncover the mechanism behind these findings.³²

Performance in the creative task is multi-dimensional, and an obvious concern is that the observed performance increase is caused by subjects increasing the quantity of ideas at the expense of their originality – a pattern that might not generally be desirable. The slider task is different in this respect as it is one-dimensional, and an overall performance increase is unambiguously positive for the principal. The Unusual Uses task allows us to look into this issue. In the Unusual Uses task, creativity is scored along three dimensions: 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). The main analysis above used an aggregate measure of these three dimensions as the dependent variable. Table 5 shows the results split up by these different dimensions. Interestingly, the pattern of results that we discussed for the overall score in our main analyses is reflected in all three dimensions of creativity. Most noteworthy, there is no evidence that the strong performance increase in the *Tournament* treatment comes at the expense of the originality of the answers. To the contrary, the share of originality points out of the total number of points – referred to as “originality rate” in the table – is even slightly higher in the *Tournament* treatment. That is, the increased originality rate suggests that the number of original uses increases more than the total number of uses.³³

It is also insightful to look at the principals’ reward decisions. 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. In the *Gift* treatment, by comparison, three quarters of the slider

³¹We find no evidence for gender differences in either the *Tournament* or the *Gift* treatment.

We can also 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 (results available upon request). 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.

³²Performance effects in response to the tournament are equal between the creative task, in which there is no reciprocity, and the simple task in which there is reciprocity. This suggests that reciprocity should not play a big role for explaining the performance increase in response to the tournament.

³³One might be concerned that statistical infrequency does not accurately reflect what is commonly understood as an “original” idea. Therefore we checked the robustness of our results on originality using subjective originality 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 and does not increase in response to the *Gift* treatment. These results are available from the authors upon request.

task principals opted for the gift, but only approximately half of the creative task principals did so. This suggests that at least some of the principals might have anticipated that the gift would not work well in the creative task.³⁴ We also asked principals to provide reasons for why they did or did not choose to implement the reward.³⁵ In both tasks the main driver for the reward decision was profit maximization, i.e. an expectation that subjects would or would not respond favorably to the treatment. Hence, the larger share of principals who opted for the gift in the slider task did so because they expected that this would raise their profits, and the share of principals who opted against the gift in the creative task did so because they did not think that the gift would pay off.

Another piece of evidence documenting the lack of reciprocity in the creative task stems from the behavior of agents when the principal refused to implement the reward. These results have to be interpreted with care as 1) the numbers of observations are low, and 2) 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.³⁶ In total, 116 agents were affected by negative reward decisions. Figure 3 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. Interestingly, the announcement that there would be no tournament has no effect on agents' performance in either task. Hence, performance responses to tournaments are similar in creative and simple tasks for both positive and negative reward decisions. Again, an asymmetry between the two tasks emerges when looking at the *Gift* treatment. While agents substantially reduce their performance in response to a negative *Gift* decision in the slider task, they remain unaffected in the creative task. Hence, neither positive nor negative reciprocity is at play in the creative task, while both are active in the slider task. 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 with other evidence showing that negative reciprocity tends to be stronger than positive reciprocity (Kube et al., 2013). Most noteworthy for our purpose is that the absence of reciprocity in the creative task holds not only for positive but also for negative reward decisions.

Overall our results suggest that tournaments effectively boost performance in both the simple,

³⁴There is no evidence for intra-session-correlation regarding the reward implementation. Results upon request.

³⁵Principals could mark whether they opted in favor of the reward 1) to maximize own profits, 2) to be nice to the agents, 3) to maximize the total payoff of all participants in their group, (4) to reward good performance (in case of the *Tournament*), or 5) other reasons. If the principal denied the reward, they could indicate whether they 1) thought that the rewards were not profitable for their own payoff, 2) did not want to provide extra earnings to the agents, 3) did not want agents to earn more than themselves, or 4) other reasons.

³⁶Principals 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.

routine task as well as in the creative task. There is no evidence for motivational crowding out in the creative task. This is good news for firms, as it suggests that creativity can effectively be raised with financial rewards. In contrast, financial gifts work well only in the routine task. Here agents reciprocate the gift in an order of magnitude similar to that found in other studies on reciprocity. There is no reciprocal response in the creative task. The absence of reciprocity in the creative task holds for both positive and negative reward decisions. Principals seem to have anticipated the lack of response because a substantially smaller fraction of employers instituted the gift in the creative task. This finding has important implications for firms as creative tasks tend to occur in settings where contracts are incomplete – that is, precisely in settings in which organizations might want to exploit reciprocal inclinations in an attempt to elicit additional effort. In the following section, we therefore further explore the absence of reciprocity in the creative task in an effort to uncover the underlying mechanism.

4 Supplementary Investigations

4.1 The Mechanism behind the Absence of Reciprocity in the Creative Task

This section discusses potential mechanisms that can explain why agents do not reciprocate the gift in the creative task but do so in the simple, routine task.

One possible explanation could be the availability of an “excuse for low performance” in the creative task. Such an excuse could be available if effort were only weakly correlated with performance in the creative task. If this were true, agents who are motivated to reciprocate by image concerns (but who don’t have an innate desire to reciprocate) might put in less effort in the creative task, hiding behind a story in which they tried hard to come up with ideas but simply did not succeed. Such an excuse is not available in the simple task, by comparison, where increased effort directly translates into more correctly positioned sliders and, in turn, increased performance. While this explanation is plausible, there is recent evidence that shows that reciprocal behavior is not affected by the availability of these kinds of “excuses” for not performing well (van der Weele et al., 2010).

A related explanation relies on the notion that the performance measure in the creative task is less sensitive to agents’ increased effort than the one in the simple task. That is, reciprocity might lead to effort increases in both tasks, but the effort increase might not translate into significantly higher output in the creative task. The results of the *Tournament* treatment speak against this explanation. In this treatment the financial stakes were identical across the two tasks, and both tasks see a performance increase of similar magnitude. This suggests that the

responsiveness of output to incentives is similar across the two tasks.

Finally, agents did not have perfect knowledge about how much surplus (in terms of the value of their ideas) they generated for the principal 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. This is true because agents' scores depended on the validity of answers as well as their originality (statistical infrequency). Hence, even though subjects probably had a rough understanding of whether they provided few or many answers and of whether their answers were off-the-chart creative or relatively unoriginal, there was still some uncertainty with respect to the exact number of points generated for the principal.³⁷ In the simple task, by comparison, the value generated for the principal was fully transparent, as subjects could easily assess how many sliders they positioned correctly. A possible explanation therefore is that this difference in agents' control over how much surplus they generate for the principal drives the differences in reciprocal behavior between the two tasks. Hennig-Schmitt et al. (2010) and Englmaier and Leider (2010) provide initial evidence that this type of uncertainty can reduce agents' reciprocal behavior. This mechanism should be more relevant in the *Gift* treatment where agents presumably want to tailor their back transfer to the perceived kindness of the gift, 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.

In order to investigate whether this mechanism explains the results, 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 separate 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*

³⁷Just like in our experiment, agents in a business context typically also have a rough sense of the value of their ideas but cannot be sure about their precise implications for the bottom line: Ideas typically go through various stages of evaluation, and there is often a relatively large time lag between the creation of an idea and its implementation. Finally, the value of an employee's idea to the firm is frequently determined by the market, say by the demand for a new product, which is also not perfectly known during the idea generation stage.

Control therefore allows us to assess whether reciprocity emerges when agents have the power to fine-tune their back transfers in the creative task.

Table 3 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.³⁸ 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 4 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 4 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.³⁹

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 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

³⁸This 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.

³⁹Performance 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.

inherent in creative tasks may hinder the emergence of reciprocity.

4.2 Analyses of Post-treatment Effects in Period 3

We now briefly turn to performance in Period 3 in order to assess whether the gift or the tournament had any long-lasting effects on performance. The analysis of Period 3 is particularly interesting for the *Tournament* treatment as any crowding out of intrinsic motivation might have been dominated by the tournament's incentive effect in Period 2 (see Bowles and Polania-Reyes, 2012 for a recent overview). Also, Period 3 performance can reveal whether winning or losing a tournament has an independent effect on subsequent performance, as subjects in the *Tournament* treatment learned at the end of Period 2 whether they did or did not belong to the best 50% (winners/losers) in Period 2 and, hence, whether or not they won the tournament prize. Period 3 endowments were identical to those in Period 1 and subjects were aware that there would be no further rewards or incentives.

Analogous to the analysis of the main treatment effects in Period 2 (see Equation 1), we assess Period 3 performance by comparing the change in treatment group performance from Period 1, with that of the control group.⁴⁰ This allows us to account for effects associated with learning and exhaustion. Table 6 shows the results.

For ease of interpretation, we separate the analyses by task. Columns I and II show the results for the slider task, and columns III and IV depict the results for the creative task. In each case, the first column (I and III) reports overall treatment effects. Interestingly, our main treatment effects carry over to Period 3 even in the absence of further gifts or tournaments. The reciprocal effect of the gift in the slider task is statistically significant and even slightly larger in Period 3 than in Period 2. Hence, reciprocity in the slider task has a persistent effect. The same pattern holds in the *Creative Transfer Gift* treatment (see columns III and IV of Table 6). As was to be expected, there is no evidence for reciprocity in Period 3 in the original *Gift* treatment (columns III and IV).

In line with standard economic theory, performance in the *Tournament* treatment is lower in Period 3 than in Period 2 due to the removal of the tournament incentive. It is noteworthy, however, that Period 3 performance exceeds Period 1 performance. This suggests that the *Tournament* has a sustainable performance-enhancing effect. Columns II and IV reveal that the tournament winners drive this overall performance increase in Period 3. They work roughly 0.5 standard deviations harder than would-be winners in the control group. Interestingly, losing a tournament does not affect subsequent behavior. Losers' performance is either not statistically significantly different from their baseline performance or is even a bit higher.

⁴⁰The results are robust to including Period 2 performance, using fixed-effects linear panel models or random-effects models.

For real-world applications, it is important to understand the mechanism behind this positive *winner effect*. One possible explanation is learning on the task. By definition, tournament winners have worked more and harder on the task than have other participants. To the extent that the tasks are subject to a steep learning curve, increased Period 2 effort and performance could translate into increased Period 3 performance even if Period 3 effort was the same for tournament winners and tournament losers. It is unlikely, however, that this effect is driven solely by learning as performance in the control group increases only slightly between Periods 1 and 2 and is relatively similar in Periods 1 and 3 (see Figure 5). Alternative explanations are that 1) the positive feedback associated with winning a tournament could heighten self-confidence and intrinsic motivation (Eisenberger and Shanock, 2003; Vansteenkiste and Deci, 2003), or that 2) winning a tournament could put individuals in a positive mood, which would in turn positively affect their performance.⁴¹ These two explanations differ with respect to whether the positive effect from winning should raise subsequent performance only for the task at hand (task-specific confidence or intrinsic motivation), or whether it should also spill over to a different, unrelated task (general mood effect).

To shed light on this issue, we conducted two supplementary tournaments. In these supplementary tournament treatments, subjects were asked to work on both the routine and the creative task in alternating orders (either slider (Period 1) - creative (Period 2) - slider (Period 3) (*SCS*) or creative (Period 1) - slider (Period 2) - creative (Period 3) (*CSC*). Identical to the main *Tournament* treatments, agents received fixed wages in Periods 1 and 3, and the principal could implement a tournament in Period 2. In total, 55 subjects participated in the *Slider-Creative-Slider (SCS)* treatment, and 46 subjects in the *Creative-Slider-Creative (CSC)* treatment. The right-hand columns of Table 3 show summary statistics for these observations. There are no significant differences in baseline performance between these two supplementary tournament treatments and the respective control groups. For this comparison, and for analyzing treatment effects below, we use the creative task control group as a benchmark for baseline performance and Period 3 performance in *CSC*, and the slider task control group to assess performance in Periods 1 and 3 in *SCS*. Note that we did not include additional control groups with varying tasks but without the tournament. We therefore cannot control for changes in Period 3 performance that are caused by the change in the tasks per se, rendering our findings below suggestive rather than conclusive. Table 7 presents the results on Period 3 performance. The results are split up by treatment and task.

Columns I and II report results from the slider task (overall and split up by winners and losers). Columns III and IV do the same for the creative task. Analogously, Period 3 treatment effects on the the mixed tasks *SCS* and *CSC* are depicted in columns V-VIII. Each regression controls

⁴¹For a review of mood effects on performance see, for instance, Lane et al. (2005). Related to mood effects, Kräkel (2008) discuss the notion that tournaments may induce a “joy of winning” which in turn affects subsequent performance. DeJarnette (2015) proposes a theory of effort momentum, with similar predictions.

for baseline performance and compares effects to the respective control group. We find no evidence for positive spillover effects – neither overall nor for winners or losers separately. The treatment effects in the mixed tasks treatments, *SCS* and *CSC* (Columns V-VIII) are positive but small and statistically insignificant. Hence, winning a tournament or receiving positive feedback may lead to higher subsequent performance, but we have suggestive evidence that such an increase is limited to the task at hand. Possible mechanisms are increased task-specific self-confidence or intrinsic motivation. Our data does not lend support for more general and task-unspecific effects on intrinsic motivation, such as mood effects.

4.3 Looking into the Effectiveness of Tournaments

Tournaments affect behavior via two different channels: 1) a concern for a good relative standing and 2) the monetary incentive (the tournament prize). For policy, it is important to understand how much of the tournament effect is driven by the (costly) prize and how much is driven by (cheap) relative performance information. To disentangle these two channels, we conducted a *Feedback* treatment that conveyed the same information about relative rank than the *Tournament* treatment but without monetary consequences.⁴²

In the *Feedback* treatment, the principal had to decide before the start of Period 2 whether or not relative rank information would be provided to her agents at the end of Period 2. The provision of relative performance feedback was costless to the principal and payoffs in this treatment were identical to those in the control group in all three rounds. When the principals opted for feedback provision, agents were informed at the beginning of Period 2 that they would learn at the end of Period 2 whether or not they belonged to the best 50% of their group. Hence, the *Feedback* treatment mirrored the information structure in the *Tournament* treatment, but without monetary consequences. We observe 56 agents with a positive feedback decision in the slider task and 68 agents with a positive feedback decision in the creative task in this treatment.

Table 8 displays the coefficient estimates of the feedback treatment from our main regressions. We find that *Feedback* increases performance by about 0.18 standard deviations in both tasks. The coefficient estimates are marginally statistically significant in period 2. The effect sizes are statistically significantly lower than those in the *Tournament* treatment in both tasks (Wald-tests, $p < 0.00$) and suggest that about one fourth of the increase in the *Tournament* treatment is driven by a concern for a good relative standing while the remainder is driven by the desire to win the tournament prize.

Effect sizes are of similar magnitude in period 3 but lose statistical significance. These overall

⁴²There is a growing literature in economics that documents the impact of relative performance feedback on behavior. So far, the evidence is mixed with respect to the direction (positive or negative) of the effect. Azmat and Iriberry (2010) and Blanes i Vidal and Nossol (2011), for example, find that feedback increases performance, whereas Barankay (2011) shows that performance rankings decrease performance.

effects hide, however, that individuals who received positive relative performance feedback perform significantly better than comparable others in the control group, mirroring our findings for tournament winners in the *Tournament* treatment.⁴³ Recipients of negative performance feedback, however, are not demotivated (in both the *Feedback* as well as the *Tournament* treatments). Their performance is not statistically significantly different from their baseline performance.

Taken together, the findings from both the *Tournament* and *Feedback* treatments suggest that the chance of receiving positive feedback enhances performance but that the effect is only strong when performing well is also rewarded with a monetary prize. This holds for both tasks as well as for ex ante incentive effects and for overall post-treatment performance.

5 Conclusion

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.⁴⁴ 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

⁴³In fact, the coefficients on *Tournament winner* and *Positive relative feedback* are statistically indistinguishable in both tasks (Wald-test creative task, 0.19; Wald-test simple task, $p=0.25$).

⁴⁴Erat and Gneezy (2015) find evidence for choking under pressure in their creative task that is more “blue sky” in nature.

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 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.⁴⁵ 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 implementation. 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

⁴⁵The 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.

more likely to give up early in the contest (for instance, Berger et al., 2013).

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.
- Alchian, A. A., Demsetz, H., 1972. Production, information costs, and economic organization. *American Economic Review* 62 (5), 777–795.
- Amabile, T., 1988. Children's artistic creativity: Detrimental effects of competition in a field setting. *Personality and Social Psychology Bulletin* 8, 573–578.
- 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.
- 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.
- Azmat, G., Iriberry, N., 2010. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics* 94 (7-8), 435–452.
- Bandiera, O., Barankay, I., Rasul, I., 2005. Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics* 120 (3), 917–962.
- Barankay, I., 2011. Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment. Working Paper.
- Bem, D. J., 1972. Self-perception theory. In: Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*. Vol. 6. Academic Press, New York, pp. 1–62.
- Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. *American Economic Review* 96 (5), 1652–1678.
- 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.

- Blanes i Vidal, J., Nossol, M., 2011. Tournaments without prizes: Evidence from personnel records. *Management Science* 57 (10), 1721–1736.
- Bonner, S. E., Hastie, R., Sprinkle, G. B., Young, S. M., 2000. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research* 12 (1), 19–64.
- 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.
- Bowles, S., Polania-Reyes, S., 2012. Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* 50 (2), 368–425.
- 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.
- Camerer, C., Hogarth, R., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19 (1), 7–42.
- Charness, G., 2004. Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics* 22 (3), 665–688.
- Charness, G., Gneezy, U., 2009. Incentives to exercise. *Econometrica* 77 (3), 909–931.
- Charness, G., Grieco, D., 2012. Individual creativity, ex-ante goals and financial incentives, mimeo.
- Deci, E. L., 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18 (1), 105–115.
- Deci, E. L., 1972. Intrinsic motivation, extrinsic reinforcement and inequity. *Journal of Personality and Social Psychology* 22(1), 113–120.
- 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.

- DeJarnette, P., 2015. Effort momentum. mimeo.
- Dickinson, 1999. An experimental examination of labor supply and work intensities. *Journal of Labor Economics* 17 (4), 638–670.
- Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review* 101 (2), 556–590.
- Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47, 268–298.
- 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., 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.
- Fahr, R., Irlenbusch, B., 2000. Fairness as a constraint on trust in reciprocity: earned property rights in a reciprocal exchange experiment. *Economics Letters* 66 (3), 275–282.
- Falk, A., Fischbacher, U., 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2), 293–314.
- Falk, A., Ichino, A., 2006. Clean evidence on peer effects. *Journal of Labor Economics* 24 (1), 39–58.
- Fang, M., Gerhart, B., 2012. Does pay for performance diminish intrinsic interest? *The International Journal of Human Resource Management* 23 (6), 1176–1196.
- Fehr, E., Falk, A., 2002. Psychological foundations of incentives. *European Economic Review* 46 (4-5), 687–724.

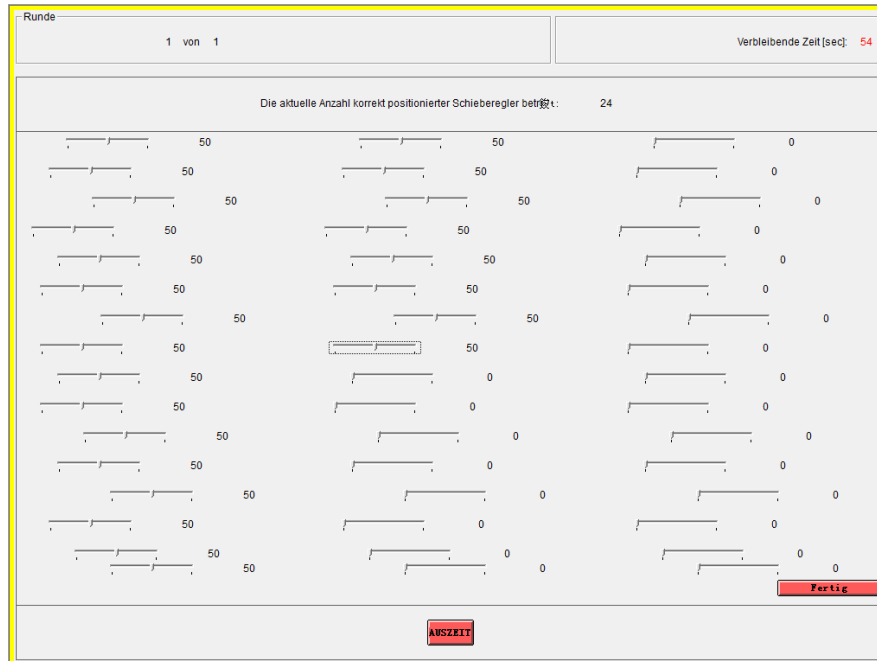
- Fehr, E., Gächter, S., 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14 (3), 159–181.
- Fehr, E., Gächter, S., 2002. Do incentive contracts undermine voluntary cooperation? IEW Working Paper No. 34.
- Fehr, E., Gächter, S., Kirchsteiger, G., 1997. 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.
- Frey, B. S., Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys* 15 (5), 589–611.
- 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.
- Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 191–209.
- 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.
- Kräkel, M., 2008. Emotions in tournaments. *Journal of Economic Behavior & Organization* 67 (1), 204–214.
- 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., Maréchal, 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.
- Lane, A. M., Whyte, G. P., Terry, P. C., Nevill, A. M., 2005. Mood, self-set goals and examination performance: The moderating effect of depressed mood. *Personality and individual differences* 39 (1), 143–153.
- 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.
- Lazear, E. P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89 (5), 841–64.
- Lepper, M. R., Greene, D., Nisbett, R. E., 1973. Undermining children's intrinsic interest with extrinsic rewards: A test of the over-justification hypothesis. *Journal of Personality and Social Psychology* 28 (1), 129–137.
- Levine, D. K., 1998. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* 1 (3), 593–622.
- 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.

- Prendergast, C., 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1), 7–63.
- 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.
- Shearer, B., 2004. Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies* 71 (2), 513–534.
- 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.
- van der Weele, J., Kulisa, J., Kosfeld, M., Friebe, G., 2010. Resisting moral wiggle room: How robust is reciprocity? IZA Discussion Paper No. 5374.
- Vansteenkiste, M., Deci, E. L., 2003. Competitively contingent rewards and intrinsic motivation: Can losers remain motivated? *Motivation and Emotion* 27 (4), 273–299.
- 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.

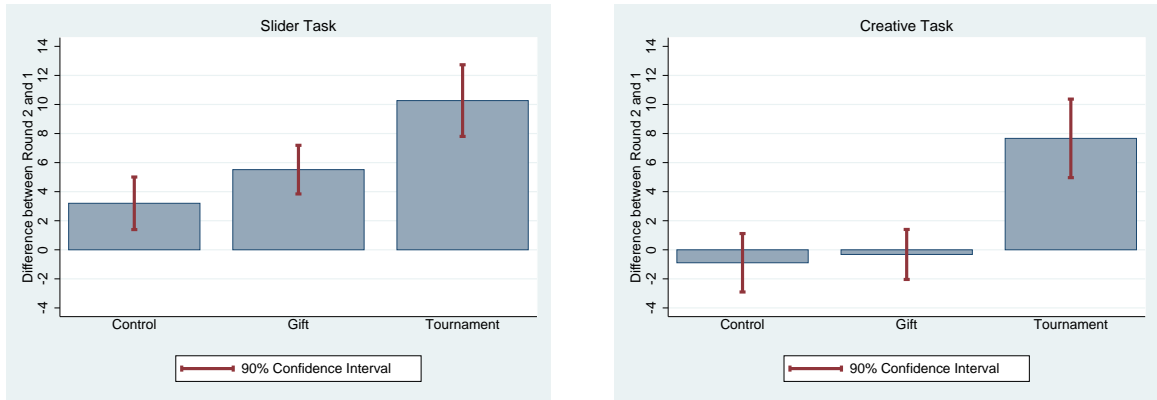
Figures and Tables

Figure 1: Screenshot of the Slider Task



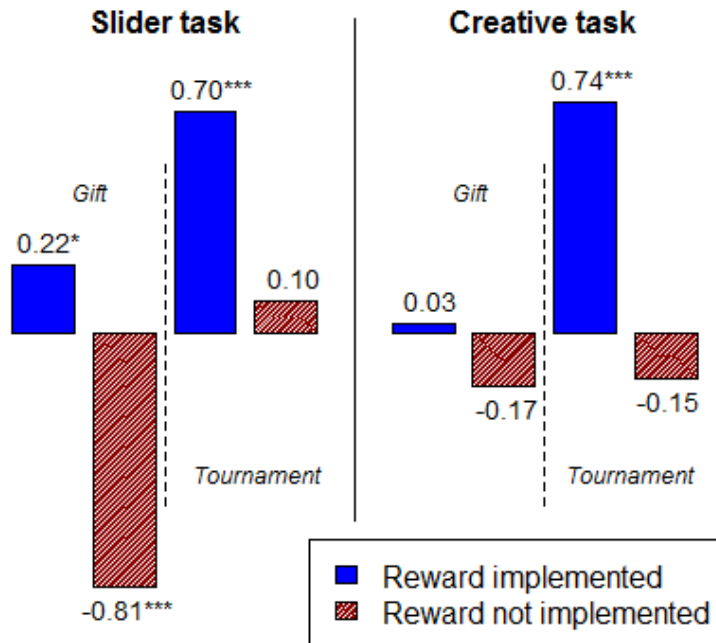
Note: The figure presents a screenshot of the slider task. The screen displays the remaining time and the number of correctly positioned sliders. The time-out button is displayed at the bottom of the screen.

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



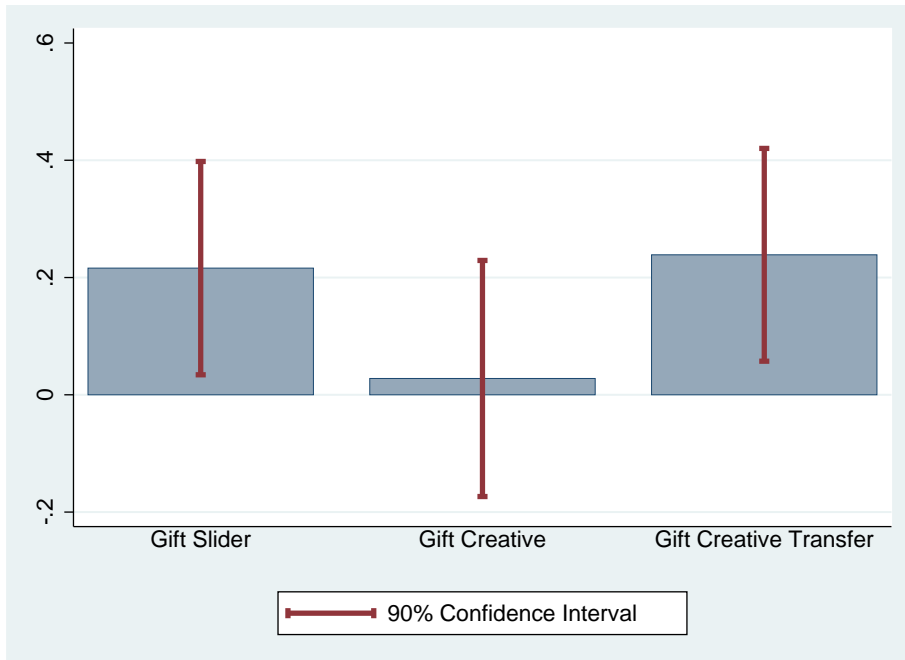
Note: The bars show the difference in mean performance between Period 2 and Period 1. The whiskers depict 90% confidence intervals of paired t-tests.

Figure 3: Effect Sizes for Positive and Negative Reward Decisions by the Principal



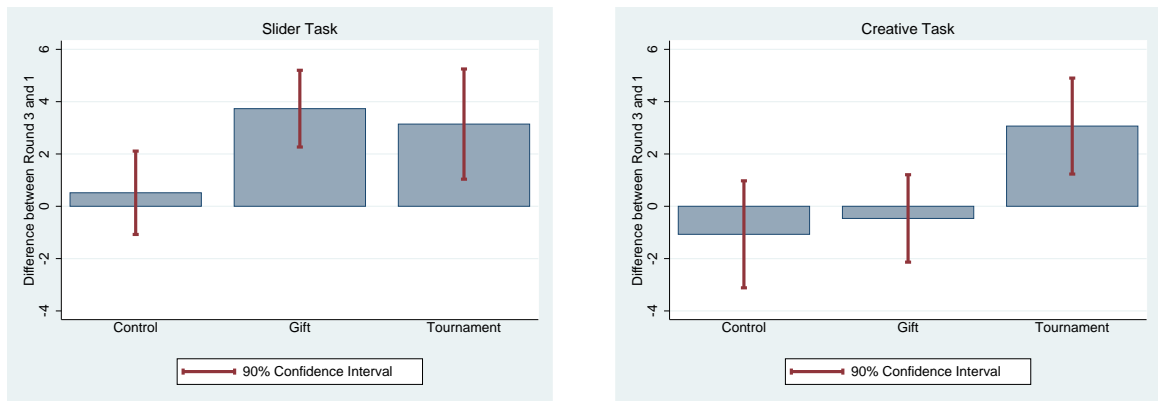
Note: The dependent variable is standardized performance of agents in Period 2. The bars show the estimated regression coefficients of separate OLS regressions. Performance is measured as the number of correctly positioned sliders in the simple task and as the score achieved in the creative task. The 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 4: 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 transferred in the creative transfer treatments. The regression controls for baseline performance in Period 1. The respective control group serves as the reference category.

Figure 5: 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 2 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: 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 3: Balance Table

	Main Treatments						Supplementary Treatments						Total
	Slider Task			Creative Task			Creative Task	Creative Transfer		Mixed Task (SCS)	Mixed Task (CSC)		
	C	G	T	C	G	T		C	G			T	
# of Subjects	75	100	95	70	135	105	90	94	94	94	96	1123	
# Agents	60	80	76	56	108	84	72	76	76	75	77	899	
# Rewarded Agents	60	60	56	56	56	60	68	68	68	55	46	712	
Mean Age	24.9	23.3	23.1*	24.0	22.8*	22.9*	23.4	23.4	23.4	22.4**	21.6***	Mean	
Share of Women	58%	43%	52%	52%	57%	53%	38%	59%	59%	58%	50%	23.1	
Economics major	50%	57%	55%	59%	61%	68%	49%	53%	53%	49%	67%*	53%	
Location												58%	
Frankfurt	67%	60%	64%	71%	71%	73%	59%	54%	54%	27%***	26%***	58%	
Mannheim	33%	40%	36%	29%	29%	27%	41%	46%	46%	22%	22%	34%	
Heidelberg	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	17.9	20.3	20.3	19.5	17.8	18.6	
	(12.5)	(14.1)	(12.8)	(11.7)	(10.7)	(10.9)	(10.3)	(11.4)	(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 4: Treatment Effects in Period 2

	I	II	III	IV
Gift Treatment	-0.019 (0.143)	0.026 (0.099)	-0.000 (0.101)	0.019 (0.100)
Gift x Slider-Task	0.375** (0.170)	0.209* (0.109)	0.195* (0.107)	0.178* (0.107)
Tournament Treatment	0.795*** (0.173)	0.736*** (0.143)	0.735*** (0.142)	0.731*** (0.142)
Tournament x Slider-Task	0.166 (0.178)	-0.058 (0.166)	-0.063 (0.167)	-0.064 (0.167)
Creative Transfer Treatment				0.087 (0.105)
Creative Transfer x Gift				0.249** (0.115)
Baseline		0.657*** (0.076)	0.631*** (0.062)	0.656*** (0.046)
Baseline x Slider-Task		0.002 (0.094)	0.100 (0.079)	0.071 (0.066)
Intercept	-0.000 (0.093)	-0.000 (0.062)	0.589 (0.496)	0.187 (0.451)
Controls	NO	NO	YES	YES
Observations	348	348	472	611
R^2	0.146	0.546	0.564	0.554

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, the creativity score in the creative task, and the amount transferred in the Creative Transfer treatments. Heteroscedastic-robust standard errors are reported in parentheses. The estimation includes data from agents in groups with a positive reward decision by the principal. 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) and observations from and controls for a supplementary feedback treatment that was run concurrently (Section 4.3 and Table 8 describe the treatment and the results). The latter are included to increase the precision with which we can estimate the control variables. 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				
	<i>Score</i> I	<i>Validity</i> II	<i>Flexibility</i> III	<i>Originality</i> IV	<i>Originality Rate</i> V
Gift	-.0220 (0.120)	0.005 (0.126)	0.002 (0.115)	-0.101 (0.144)	-0.0180 (0.023)
Tournament	0.746*** (0.161)	0.805*** (0.169)	0.601*** (0.148)	0.719*** (0.181)	0.045** (0.021)
Period 1	0.615*** (0.061)	0.598*** (0.062)	0.613*** (0.056)	0.523*** (0.071)	0.173** (0.071)
Intercept	1.908 (2.029)	1.620 (1.917)	1.233 (1.685)	3.168 (2.72)	0.293 (0.395)
Controls	YES	YES	YES	YES	YES
Observations	240	240	240	240	215
R^2	0.484	0.470	0.515	0.346	0.109

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 to IV report the results on the different sub-dimensions of the creativity score. Column V displays treatment effects on the originality rate. The originality rate equals achieved originality points divided by the total number of points for a subject's answers (subjects with zero points were dropped in this column). The estimation includes data from agents in groups with a positive reward decision by the principal. 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) and observations from a supplementary feedback treatment (with separate treatment dummies, see Section 4.3 and Table 8). The latter are included to increase the precision with which we can estimate the control variables. Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Treatment Effects in Period 3

	Slider		Creative	
	I	II	III	IV
Gift	0.259** (0.110)	0.268** (0.110)	0.033 (0.113)	0.031 (0.114)
Tournament	0.223 (0.153)		0.365*** (0.126)	
Creative Transfer			-0.038 (0.122)	-0.035 (0.122)
Creative Transfer x Gift			0.366*** (0.110)	0.362*** (0.110)
Tournament Winner		0.503*** (0.189)		0.493*** (0.159)
Tournament Loser		-0.018 (0.159)		0.242* (0.140)
Controls	YES	YES	YES	YES
Baseline	YES	YES	YES	YES
Intercept	YES	YES	YES	YES
N	232	232	379	379
R^2	0.712	0.728	0.540	0.548

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. Heteroscedastic-robust standard errors are reported in parentheses. The estimation includes data from agents in groups with a positive reward decision by the principal and observations from a supplementary feedback treatment (see Section 4.3). The latter are included to increase the precision with which we can estimate the control variables. 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). Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Post-treatment Effects of the Tournament in Period 3 by Task

	Slider Task			Creative Task			Slider-Creative-Slider (SCS)			Mixed Tasks Creative-Slider-Creative (CSC)		
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	
Tournament	0.245* (0.137)		0.340*** (0.126)		0.111 (0.109)				0.055 (0.121)			
Tournament Winner		0.534*** (0.180)		0.449*** (0.170)		0.064 (0.122)				0.091 (0.146)		
Tournament Loser		-0.030 (0.146)		0.236* (0.134)		0.158 (0.142)				0.007 (0.132)		
Standardized Performance in Period 1	0.888*** (0.060)	0.848*** (0.063)	0.693*** (0.080)	0.670*** (0.086)	0.888*** (0.052)	0.891*** (0.053)	0.648*** (0.089)				0.648*** (0.090)	
Intercept	0.000 (0.073)	0.000 (0.073)	0.000 (0.093)	0.000 (0.094)	0.000 (0.073)	0.000 (0.073)	0.000 (0.094)				0.000 (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. Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The estimation includes data from agents in groups with a positive reward decision by the principal.

Table 8: Period 2 and Period 3 Effects of the Feedback Treatment

	Slider-Task	Creative-Task
Feedback Period 2	0.184* (0.108)	0.182* (0.102)
Feedback Period 3	0.118 (0.132)	0.171 (0.129)
Positive Relative Feedback Period 3	0.303 (0.196)	0.319** (0.148)
Negative Relative Feedback Period 3	-0.028 (0.135)	0.035 (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 4 while post-treatment effects for Period 3 are from Table 6. Heteroscedasticity-robust standard errors are reported in parentheses. 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. 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.⁴⁶ 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.]

⁴⁶For 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.