Rewards and Performance: A Comparison Across a Creative and a Routine Task *

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

Preliminary draft, January 2015

Abstract

Motivating employees to be creative is an important challenge for many firms. A large literature documents how employee performance can be fostered by explicit and implicit rewards in simple routine tasks. This paper investigates whether rewards are similarly effective in enhancing creative performance as a substantial psychological literature suggests that rewards may undermine intrinsic motivation, and thus lower creative output. We report results from a real-effort lab experiment with more than 1000 subjects that studies the impact of two types of monetary rewards, a performance-dependent tournament prize and a performance-independent wage gift, and their impact on creative as well as purely effort-based, routine performance. The design allows us to study whether the effectiveness of monetary rewards depend on the nature of the task and the performance contingency of the reward. Our results show that routine as well as creative task performance significantly increase in response to the tournament incentive. A wage gift triggers reciprocity of subjects in the routine task but not in the creative task. We explored this with additional treatments and find that reciprocal behavior emerges in the creative task if subjects receive information on the value of their ideas and can control the transfer of benefits to the principal. This suggests that agents reciprocity is sensitive to the transparency of principal's surplus. Important implications for human resource management are discussed.

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

---

*We gratefully acknowledge financial support by the framework program "Economics of Science" by the Federal Ministry of Education and Research, Germany. Further, we thank for financial support provided by the SEEK research programme.

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

[‡]Erasmus University Rotterdam 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 on non-routine tasks that require creative thinking has strongly increased during the last decades (Autor et al., 2003; Florida, 2002). For organizations, this poses the challenge how to motivate employees to make full use of their creative potential.

Extensive empirical evidence has demonstrated that financial rewards can significantly enhance work motivation and performance of employees. However, a substantial number of studies show that the impact of rewards may fundamentally differ when tasks involve a high level of intrinsic motivation as assumed to be particularly pronounced for employees in creative jobs. Specifically, evidence has been provided which suggests that extrinsic rewards may crowd-out intrinsic motivation and thereby lower performance. Based on this concern, many scholars have argued that motivating creative output of employees is fundamentally different from motivating routine performance (for instance, Amabile, 1996; Shalley et al., 2004; Byron and Khazanchi, 2012).

The present study investigates whether creative performance can be fostered by explicit and implicit financial rewards. We conducted a large-scale experiment which allows (1) to compare the effectiveness of two main types of monetary rewards, a performance-dependent reward vs. a performance-independent wage gift, and (2) to study their impact on performance in a non-routine creative task as well as in a simple routine task. Thus, we are able to directly compare how explicit performance rewards as well as implicit rewards, such as wage gifts, affect creative vs. routine task performance in one experimental set-up.

There is a long tradition in economics to investigate the impact of work incentives on motivation and productivity in simple and routine jobs (see Prendergast, 1999, for an overview). In the field, researchers observed work performance, for instance, when 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. Most studies on routine tasks confirm what standard economic theory predicts: financial incentives have a positive effect on performance because agents increase effort as long as their benefits from additional output exceed their effort costs (Holmstrom and Milgrom, 1991).[1]

However, explicit financial incentives are not the only way to trigger workers' performance. An established literature on gift exchange documents that people reciprocate gifts that are assigned

[1]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 (e.g., Lazear, 2000) or tournaments, which reward workers on the basis of their relative performance (Harbring and Irlenbusch, 2003).

independently of performance, such as above market clearing wages (see Fehr and Gächter, 2000, for an overview). A myriad of laboratory experiments shows that subjects increase chosen effort levels subsequent to receiving a monetary gift (for instance, Fehr et al., 1997). More recently, the positive wage-effort-relationship has also been demonstrated in real-effort tasks in the laboratory, such as in solving math equations (for instance, Brueggen and Strobel, 2007), stuffing envelopes (for instance, Hennig-Schmitt et al., 2010), coding tasks (for instance, Englmaier and Leider, 2010) or real work tasks in the field such as tree planting (Bellemare and Shearer, 2009).

Taken together, these findings inform human resource management on how to optimally reward employees in jobs that involve a clearly defined and repetitive workflow. Yet, it is critical to understand whether these insights into the effectiveness of different rewards also hold for jobs which rely on creative performance.[2] So far, economic theory does not predict that the effectiveness of rewards depends on the nature of the task. However, studies which review the existing evidence on the impact of explicit financial incentives on task performance show that a positive relationship is less likely to be observed for tasks which are cognitively more complex, such as creative problem-solving tasks (Bonner et al., 2000; Camerer and Hogarth, 1999). A controversially debated reason for why financial incentives sometimes fail to increase performance is the presence of a high intrinsic motivation on the task.[3] A substantial number of studies demonstrated that financial incentives can even harm performance by causing a crowding-out of intrinsic motivation (Deci et al., 1999). A large body of this literature can be found for creative tasks as they are a very prominent type of task where intrinsic motivation is assumed to be vital and strongly attached to the task (for an overview, see Amabile, 1996; Shalley et al., 2004).

Several psychological theories offer explanations for why individuals' creativity may suffer if exposed to external motivators. According to self-perception theory (Bem, 1972), a reward causes a shift in peoples' perception about 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 the reward is removed again, individuals are less motivated to perform on the task. This "overjustification hypothesis" has been confirmed in several studies (for instance, Deci, 1971; Lepper et al., 1973). Based on these findings, Edward Deci and Richard Ryan (1980, 1985) developed cognitive evaluation theory, which posits that intrinsic motivation is determined by the psychological needs for

[2]According 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 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.

[3]Other reasons which have been discussed are, for instance, that incentives may leave performance unaffected or even harm performance if they cause people to overlearn a heuristic, to exert "too much effort" when a low effort habit would suffice or when arousal caused by incentives raises self-consciousness behind an optimal level (Camerer and Hogarth, 1999).

autonomy and competence. Subsequently, a substantial number of studies emerged that show negative effects of rewards on intrinsic motivation and argue that the controlling aspects of rewards decrease the feeling of autonomy, and thus intrinsic motivation (Amabile, 1988, 1996; Joussemet and Koestner, 1999; Deci et al., 1999). In contrast, rewards have also be shown to positively affect intrinsic motivation and creative output which has been related to their informational aspects. In particular, performance-dependent rewards were discussed to enhance feelings of self-determination as they signal the lack of control of the principal on agents effort, and satisfy individuals' need for competence if assigned for excellent performance (Eisenberger et al., 1999; Eisenberger and Rhoades, 2001).

Still, "despite the many studies examining rewards and creativity in the psychological, educational and organizational literature, rewards' effects on creativity – and the factors that serve to mitigate or enhance those effects – remain unclear." (Byron and Khazanchi, 2012, p.1). Most notably, it remains an open question whether the types of rewards which are of economic relevance in many principal-agent relationships, such as tournament bonuses or wage gifts, provoke undermining effects on intrinsic motivation in creative tasks.

The present study investigates the impact on creative performance if principals provide a wage gift as an unconditional performance reward or a tournament incentive as an explicit reward which is made conditional on performance. We contribute to the economic literature where crowding-out effects have so far been studied mainly in the context of intrinsically motivated activities other than creative tasks (see, for instance, Frey and Jegen, 2001 or Gneezy et al., 2011, for an overview) and their prevalence is controversially debated (for instance, Fang and Gerhart, 2012; Fehr and Falk, 2002; Charness and Gneezy, 2009). Some exceptional economic studies on the effectiveness of rewards on creativity either focused on how effects differ depending on the size of the reward (Ariely et al., 2009) or the type of creative task (Charness and Grieco, 2012).[4] We extend the nascent economic literature on creativity by investigating to which extent previous findings for the effectiveness of monetary rewards in routine tasks can be transferred to creative tasks. Moreover, Fehr and Falk (2002) stated a general need to study how intrinsic motivation interacts with implicit incentives such as wage gifts, as previous economic studies on the crowding-out effect have only examined the interaction between intrinsic motivation and explicit rewards which are conditional on task performance. Our study contributes to this existing literature on crowding-out effects by directly comparing effects of performance-dependent vs. performance-independent rewards on creativity. Any differences between reward groups and a control group can be causally attributed

[4] A further exception in the economic literature is Eckartz et al. (2012). They implemented a creative task, and for comparison two control tasks (Raven's IQ and a number-adding task) in an experimental set-up. They find that neither a tournament reward nor a piece rate scheme had an effect on any task, which makes it hard to draw clear conclusions on whether the rewards did not enhance creative performance or other limiting factors, such as a high baseline motivation of subjects may hampered performance effects in their study.

to the type of reward and comparisons across tasks allow to causally attribute differences in the effect to the type of task. Further, a three stage design allow us to observe performance effects of individuals before, during and after the presence of a reward. This is particularly interesting for the performance-dependent reward to study whether intrinsic motivation is affected only after explicit incentives are withdrawn again.

We conducted a laboratory experiment with more than 1000 subjects who were randomly assigned the role of a principal ("employer") or an agent ("employee"). In the experiment, four agents worked for three periods for one principal on either a creative or a routine task. In the basic set-up, the payoff of principals depended on the work effort by their four agents who received an exogenously given fixed wage in each period. The first period provides a measure of an individual's baseline performance comprising agents' ability, intrinsic motivation to perform on the task and other-regarding preferences towards the principal. After period 1, principals could decide whether they would like to assign additional rewards to their agents (at own expenses). In the *Tournament* treatment, the principal could opt to institute an additional monetary prize to the 50% best performing agents in his group. In the *Gift* treatment, the principal could opt for a monetary gift, half as high as the tournament prize, to all agents in his group. Both reward schemes had identical costs to the principal and identical benefits to the group of workers; they differed only in the distribution of benefits among workers. Thus, we can evaluate which of the two schemes is preferable given a fixed budget for rewards. Whether or not the principal opted for or against the rewards was announced right before the start of period 2. This second period, therefore, allows us to observe whether the performance-independent gift or the performance-dependent tournament reward leads to an increase in effort of the agents and whether or not the effort response depends on the nature of the task. After period 2, winners and losers of the tournament were privately revealed. Then, subjects were asked to work for the principal for one more period without additional rewards (identical to period 1) in order to investigate whether the treatment effect of the gift is lasting and if tournament participants show a motivational response to the information of being a winner or loser in the competition. In addition to the two monetary reward treatments, we ran a control group for each task in order to be able to take learning and exhaustion effects into account. Moreover, an additional *Feedback* treatment allows us to disentangle the effects of the monetary reward from the effect of the relative performance information per se, that is inherent in the *Tournament* treatment. In this treatment, the principal could for informing agents whether they belong to the two best performing agents of their group or not. This information was free of cost for the principal and was not associated with a material consequence for the agents.

We find a substantial and positive incentive effect of the *Tournament* in both tasks (routine and creative) in the second period. About one third of this performance increase is driven by a concern for good relative standing as is revealed in the *Feedback* treatment. Subsequent to

learning whether they were winners or losers at the end of period 2, winners in the *Tournament* perform better than the control group in period 3 even though there is no further tournament that they compete in. This evidence is inconsistent with the notion that financial rewards crowd-out intrinsic motivation, and thus creative performance. To the contrary, it suggests that a performance-dependent rewards in the form of a tournament incentive increases creative performance and has even long-lasting effects on intrinsic motivation of tournament winners. Two supplementary tournament treatments with mixed tasks (creative and routine in alternating orders) show that this increase in intrinsic motivation is task-specific and does not spill-over to a different task. Interestingly, our results show an asymmetry of agents' responses to the performance-independent gift 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 exchanges (see, for instance, Fehr and Gächter, 2002). However, agents do not reciprocate the gift in the creative task. This finding is surprising and suggests an asymmetry in behavioral responses to performance-independent rewards between routine and creative tasks.

In order to investigate this asymmetry, we conducted two further supplementary treatments with the creative task. A potential reason for the contrary effect is that the agent's generated surplus for the principal is easily observable in the routine task but intransparent in the creative task. In the latter, agents were asked to submit creative ideas without knowing the value of their ideas. This is typically the case in creative tasks in practice as the value of ideas is not known until they are evaluated or implemented. In the supplementary treatments, agents are asked to perform on the same creative task, but receive feedback after each period how much surplus they generated by their ideas for the principal. Subsequently, agents could choose how much of the generated surplus they actually want to transfer to the principal whereas non-transferred surplus is lost. In order to investigate whether the introduction of surplus information and transfer possibility triggers reciprocity, we implemented one treatment group where principals could assign a gift in period 2 and a control group with surplus information and transfer decision but without reward. The results demonstrate that the introduction of transparency on the value of agents' effort to the principal evokes reciprocal behavior. This finding suggests that wage gifts for employees in creative tasks may not be efficient in inducing reciprocal behavior if the information on principals surplus is not transparent. This is in line with evidence from previous studies showing that transparency on payoffs is important to a positive wage-effort-relationship (Hennig-Schmitt et al., 2010).

The paper is structured as follows. Section 2 describes the experimental set-up, the tasks and treatments. We briefly summarize our hypotheses in Section 3 and present our results in Section 4. Section 5 concludes.

2 The Experiment

Does the response to rewards systematically differ in creative compared to simple routine tasks? And given a fixed budget, which type of reward is preferable to foster creativity: a bonus for the most creative employees or a monetary gift to all employees which may induce reciprocal behavior? We designed an experiment that addresses these research questions as well as the underlying theoretical mechanisms. In the following, we introduce the experimental tasks, the set-up and treatments as well as 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 simple routine task (Gill and Prowse, 2012) which requires real-effort of agents. The slider task 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 consists of a computer screen displaying 48 sliders on scales that range from 0 to 100. 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. Each slider can be adjusted and re-adjusted an unlimited number of times. While moving the mouse, the subject cannot be sure whether they positioned the slider at exactly 50. The exact position of the slider is displayed to the right of the scale as soon as the subject stops using the mouse. The number of correctly positioned sliders at the end of the allotted time corresponds closely to the effort exerted by the subject, and therefore measures subject's performance.[5] Before period 1 started, subjects were given one minute to practice the task.

In order to measure creative performance, we implemented the "Unusual Uses Task". Originally developed as Guilford's Alternative Uses Task (Guilford, 1967), it was later on also incorporated in the Torrance Test of Creative Thinking ("TTCT", Torrance, 1968).[6] The TTCT is the most widely used test to assess an individual's capacity for creativity, and its validity has been confirmed in a large number of studies (Kim, 2006). In the unusual task, participants are asked to name as many, as different, and as unusual uses for an ordinary item, such as a tin can. The unusual uses task captures a central element in applied business innovations: a recombination

[5] Keyboards were disconnected during the task to prevent the usage of the arrow keys. Figure 1 in the Appendix shows an example of the screen as it was presented in the experiment.

[6] In order to assess overall creative potential, the TCTT also comprises a number of figural elements, which require drawing skills. Unusual uses, however, best presents the type of creativity that we want to study.

of existing bits of knowledge in novel ways (Weitzman, 1998; Simonton, 2004).[7] Specifically, the task requires divergent thinking, known as "out-of-the-box" thinking, which is among the most important components of the creative process (Runco, 1991). In order to test whether the creative task is more intrinsically motivating than the routine task, we asked about 100 subjects to rate their interest in the tasks on a Likert-scale 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).

In the experiment, subjects had to sequentially work on three different items: a sheet of paper, a tin can, and a cord. Subjects were informed that they should not limit themselves to one particular size of the item. Moreover, the unusual use can require more than one of the items, for instance, the use can 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.[8]

In the creative task, just as in the slider task, subjects had a test period of one minute. In this test period, the item "old tire" was given to familiarize with the task and the entry mask on the screen. We used the three standard measures to evaluate subjects' responses: fluency, flexibility, and originality (Guilford, 1959) and informed subjects about the evaluation procedure.[9] 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 beyond all possible reality are not counted. Examples for a valid use of a tin can are, for instance, a flower pot, a pen container or a drum. In contrast, examples for invalid answers are to use a tin can as a television, a computer or a window.[10] In the experiment, each valid use is scored with 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, a candleholder falls into the category decoration; a baking-tin into the category molds, while a rattle or a drum are music instruments. For each category which is mentioned, the subject receives one point in the experiment. Frequent categories for the tin can are "non-food containers" (for instance, a pen container), "sport devices" (for instance, a football), and

[7]Different types of creativity economists have looked at are, for instance, word formation from a given letter set Eckartz et al. (2012), open problem solutions (Charness and Grieco, 2012) or exploration in a computer simulation game (Ederer and Manso, 2012).

[8]Controlling 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 group.

[9]The original Guilford Test uses a fourth criterion for scoring, elaboration, which refers to the detailedness 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.

[10]Usual uses, such as storing food in case of the tin can, were also 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. That is why we also scored usual uses as valid answers. Excluding usual uses does not alter our results (results available upon request).

"communication" (for instance, a tin can phone).[11]

Finally, the originality of responses was measured by the statistical infrequency of answers. Examples for an 'original' use of a tin can are an insect trap, a bedstead or an animal house. 'Very original' answers include using it as a scarecrow, a shower head, a treasure chest or a grill. Table 1 illustrates further examples of frequent answers and categories as well as original, very original, and invalid example responses for all items.

In order to get an idea of the frequency of responses, we conducted a pre-test with 127 participants that worked on the three items under a fixed wage scheme. We tabulated all valid answers of an item according to their frequency and constructed a rating scale to assess answers in the experiment. This scale allotted one point to an answer if it belonged to the 8% most rare answers in the pre-test ("original") and two points to answers that belonged to the 1% most rare answers in the pre-test ("very original"). Scoring was conducted by research assistants who were carefully acquainted with the scoring procedures and blind to the treatments. In comparison to expert ratings, evaluating originality in this manner is more objective due to the statistical approach. However, some scope for subjectivity remains in the evaluation of the unusual uses task, too, as this is a typical feature for creative tasks.

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) voluntary financial transfers from the principal to the agent, and 2) agents' effort to affect the principal's payoff. 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 periods à 3 minutes, depending on the treatment either on the routine or the creative task. Employees received a fixed wage for each period which was exogenously given by the experimenter and announced before each period. Depending on the treatment, principals could assign a reward before period 2 which agents receive on top of their fixed wage. In period 3, agents received again a fixed wage without additional rewards. The three-period design allows us to measure baseline performance of agents under fixed wage in period 1,[12] the performance response to the reward in period 2, as well as ex-post performance under a fixed

[11] The number of categories is comparable between items and adds up to approximately 55 per item.

[12] As 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, the social preference towards the principal as well as a subject's ability.

wage in period 3. Detrimental effects of the rewards on intrinsic motivation cannot necessarily be detected when rewards are present as they might be outbalanced by positive incentive or reciprocity effects of the rewards. Therefore, we observe performance in period 3 to measure potential long-term effects of rewards on intrinsic motivation. Employers payoff consisted of a fixed wage component and a variable amount determined by the total performance of the four agents in their group. All payoffs assigned during the experiment were stated in "Taler" as the experimental currency unit.[13] In the routine task, principals received 5 Taler for each slider 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) as well as 5 Taler for each originality point, hence, 5 Taler per original answer and 10 Taler per very original answer given by their four agents. Agents were informed about the scoring procedure in detail in the instructions.

In order to implement opportunity costs 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 agents clicked the time-out button, the computer screen was locked for 20 seconds, prohibiting the entry of creative ideas or the movement of sliders, and 5 Taler were added to the agent's payoff. This 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 and due to the absence of alternative activities (Eckartz et al., 2012; Mohnen et al., 2008).[14]

2.3 Implementation of Treatments

In order to address the research questions raised above, we implemented the possibility for the principal to offer a costly reward before period 2. In the *Tournament* treatment, principals were able to announce a performance reward for the top 50% performers in the work group.[15] In the *Gift* treatment, the principal could assign a monetary gift to all employees in the group. The two reward treatments are complemented by a control group and a *Feedback* treatment without financial reward as comparison groups. Thus, the experiment has a 4x2 design, that is, four treatments were conducted for each of the two tasks. Independent of principal's decision on the reward, agents were informed about the principal's possibility to administer an additional reward (or feedback) as well as its cost and benefits. In the financial reward treatments, both

[13]The exchange rate was 100 Taler = 1 Euro.

[14]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 60 seconds trial period in which subjects could test the time-out button.

[15]Tournaments are a widespread form of performance rewards for creative performance and innovations in practice (Brunt et al., 2012; Kremer and Williams, 2010). For instance, companies increasingly allocate creative tasks to online platforms with creative contests (such as Innocentive or Jovoto) 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).

principals and agents received a fixed wage of 300 Taler at the beginning of each period. In the *Tournament* and in the *Gift* treatment, principals were given the choice, before period 2 started, to allocate rewards worth 1200 Taler to their agents, at a cost of 200 Taler. This results in an efficiency factor of 6, which renders the reward relatively attractive.[16]

Subjects were instructed that principals did not receive any information on the performance of their agents until the very end of the experiment. Thus, agents knew that the reward decision was independent from their performance in period 1. This avoids an endogenous selection of agents who receive the reward.

The reward sizes in the *Tournament* and the *Gift* treatment were chosen such that the cost of their provision for the principal as well as expected payoffs of the agents were identical across these two treatments. However, the treatments differed in the way the rewards were allocated to the agents. In the *Gift* treatment, all four agents of a group received an immediate gift of 300 Taler if their principal opted for the reward implementation.

In the *Tournament* treatment, the same amount of 1200 Taler was at stake, but payment was linked to the performance of the agents. In case the principal opted for the reward, it was announced at the start of period 2 that the best-performing 50% of the agents in this period (hence, two out of four) would receive a tournament bonus of 600 Taler at the end of period 2. Subjects were also told that the evaluation of performance was done immediately by the experimenter based on the criteria detailed in the written instructions. Between periods 2 and 3, subjects in the *Tournament* treatment were informed whether they won the reward or not. Prior to period 3, subjects in all treatments were informed that there would be no further rewards or feedback and that endowments and payment structure were identical to period 1.

In the control group and the *Feedback* treatment the fixed wage was also 300 Taler in periods 1 and 3. In period 2, the fixed wages in these two treatments mirrored those in the monetary reward treatments in case the reward was implemented, that is 100 Taler for the principal and 600 Taler for each agent (given that agents have an expected payoff of 300 Taler in the *Tournament* treatment). This ensures that any performance differences in the monetary treatments compared to the control group or the *Feedback* treatment are solely driven by the rewards and not by other motivations such as distributional concerns or income effects. Inequality aversion and concerns about relative payoffs have been shown to significantly influence subjects' behavior (Fehr and Schmidt, 1999) and should be separated from pure intention-based reciprocity, which goes beyond distributional concerns (Charness, 2004). Table 2 provides an overview of the fixed and variable pay components for all periods, treatments, and roles.

[16]This is important in our setting as we are mainly interested in agents' response to a positive reward decision rather than whether or not principals opt for the reward. Multiplying monetary transfers by an efficiency factor like the one we use is a standard procedure in gift exchange games (see, for instance, Brandts and Charness, 2004).

The *Feedback* treatment was identical to the *Tournament* apart from the monetary consequences. In this treatment, the principal had to decide before the start of period 2 whether or not relative performance information should be provided to the agents of their group at the end of period 2. The implementation of relative performance feedback was costless to the principal. In case they opted for feedback provision, agents were informed between periods 2 and 3 whether or not they belonged to the best 50% of their group. This treatment was designed to disentangle the reward from the feedback effect in the *Tournament* treatment.[17]

In this study, we are mainly interested in agents' response to the implementation of the reward and fixed payoffs as well as reward sizes are chosen such that a comparison of treatments and the control group is possible. If the principal decided against the reward, agents and principal received 300 Taler as fixed endowment for period 2 in the *Gift* and *Tournament* treatment. In this case, period 2 endowments differ between the monetary reward treatments on the one hand, and the control group and *Feedback* treatment on the other hand. Therefore, we are not able to directly compare the performance effects of agents in sessions where the reward was not implemented by the decision of the principal. However, we will present some descriptive statistics on the frequency of reward implementations across treatments and tasks as this provides some interesting insights to which extent principals already anticipate the efficiency of the different rewards and to which extent this may differ across tasks. Further, we elicit some information why principals decided in favor of or against the reward.

After period 3 ended, agents were confronted with a couple of brief decision tasks which included incentivized control measures for subjects' risk and tournament aversion and social preferences. Moreover, we included some questions about subjects' socio-demographic characteristics such as their gender, field of study, education and high school grade, and leisure activities as well as questions regarding their personality traits including the BigFive.[18]

[17] As described in the introduction, feedback itself might have an effect on the intrinsic motivation of agents as it is supposed to affect agents' perceived competence and self-determination (Eisenberger et al., 1999).

[18] Risk aversion was elicited with a simple multiple price list design which allows to infer subjects' certainty equivalent for a given lottery (see, for instance, Bradler, 2009). Tournament aversion was observed with a 3-minute counting task (see Abeler et al., 2011) where subjects could choose between a piece rate and a tournament reward scheme. Additionally, we implemented a measure for the social preference type (Murphy et al., 2011) as well as people's reciprocal inclinations in a simple gift-exchange game similar to Englmaier and Leider (2010). For each incentivized control measure subjects were informed that they interacted with a different randomly matched subject. At the end of the experiment, one of these four incentivized measures was randomly chosen for payoff. The personality scales that we used were the Big Five and reciprocity questions from the Socio-Economic Panel (see Gerlitz and Schupp, 2005) as well as the Gough scale, which is a validated scale for creative personality measured by a 30-items adjective check list (Gough, 1979).

2.4 Procedures

The experiment was conducted at the experimental laboratories of 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. Written instructions informed them about the random matching of groups and roles ("employer" or "employee"), the basic structure of the experiment, the work task (routine or creative depending on the treatment) and the scoring of the task. A translation of the original instructions can be found in the Appendix A. Further information on subsequent fixed wages in periods 2 and 3, as well as on the assignment of rewards were presented on the computer screen. Before the experiment started, subjects had to complete control questions on how actions determine payoffs.
As principals' active part in the experiment was limited to the reward decision, they were offered to work on the respective task as well in order to avoid boredom. However, they were informed that their payoffs were not influenced by their own performance. A session lasted about 75 minutes on average. For payment, the subjects' payoff in the experimental currency unit "Taler" was converted into Euros at an exchange rate of 100 Taler = 1 Euro. Average earnings were approximately 15 Euros.

3 Hypotheses

For our empirical analyses, we derive hypotheses from established economic theories. However, we will analyze and discuss to which extent our experimental results are line with theories on crowding-out and crowding-in effects on intrinsic motivation which have been controversially discussed in the economic as well as psychological literature.

According to standard principal-agent-theory, financial incentives, such as tournaments for a monetary prize, have a relative price effect. That is, they increase the marginal benefit of exerting effort when there is a direct link between effort and performance, and thus, should increase performance. For the prediction of a positive incentive effect of a tournament prize on performance, no differentiation is made between tasks of different nature (Alchian and Demsetz, 1972; Lazear and Rosen, 1981):

H1a: ***The tournament incentive increases performance in period 2. The magnitude of the effect is similar in the creative as in the simple routine task.***

The effect of performance-dependent rewards on creative output was controversially debated

in the psychological literature. One camp in the literature argues that conditional rewards have informational aspects that increase intrinsic motivation and thereby, creative performance. Several channels have been discussed for this relationship. First, perceived autonomy increases as the implementation of an explicit, performance-based reward signals to agents that the principal has little control over what the agents do in the absence of such a reward (Eisenberger et al., 1999; Eisenberger and Cameron, 1996). Secondly, if rewards are assigned conditional for excellent creative performance, they increase agents' feelings of competence. And thirdly, they signal to the agent that the task is of particular importance for the principal. We will refer to this approach as a "crowding-in" of intrinsic motivation. A positive performance response in period 2 can therefore reflect both, an incentive effect and/or a crowding-in effect of intrinsic motivation.

Another camp in the psychological literature argues that conditional rewards may predominantly be perceived as controlling agents' self-determination which may cause a "crowding-out effect" on intrinsic motivation (for instance, Deci, 1972; Lepper et al., 1973; Deci et al., 1999). However, it depends on the magnitude of the incentive effect relative to the crowding-out effect if the overall effect on intrinsic motivation is positive or negative. A detrimental effect would only occur when crowding-out effects overcompensate the incentive effect (see Bowles and Polania-Reyes, 2012 for a recent overview). In contrast, a positive performance effect in period 2 does not necessarily mean that no crowding-out effect occurred but it may be dominated by a strong incentive effect. In order to further disentangle a potential crowding-out or crowding-in effect, we study performance effects in period 3 where the tournament incentive is withdrawn again and lasting effects on intrinsic motivation become apparent.[19]

In the absence of learning or fatigue, that we control for via the control group, standard economic theory predicts that performance returns to baseline level in period 3, which is predicted to be zero for egoistic agents. Thus, the *Tournament* should not reveal a significant difference between performance in period 3 and 1 (compared to the control group):

H1b: ***The tournament has no significant ex-post effect on performance in period 3, neither in the creative nor in the simple routine task.***

In contrast, if performance-depended rewards cause a crowding-in effect, performance in period

[19]Further, we isolate the *Tournament* effect from the effect of an announcement of performance evaluation by comparing the *Tournament* with the *Feedback* treatment. Standard economic theory assumes that the announcement of relative performance feedback without monetary consequence should have no effect on an individuals' performance in both tasks. Thus, performance in the *Tournament* treatment should be larger than in the *Feedback* treatment in period 2. In contrast, several studies in the psychological literature show that the mere announcement of performance evaluation can reduce creative performance (Amabile, 1979; Deci et al., 1999) due to its controlling aspects. However, it has been argued that a monetary reward attached to a performance evaluation is assumed to sharpen the undermining effect. Thus, we expect that if crowding-out effects occur, they should be more pronounced in the *Tournament* than in the *Feedback* treatment.

3 should remain higher than baseline due to heightened intrinsic motivation. Vice versa, if performance-depended rewards cause a crowding-out effect, performance in period 3 should decrease below the baseline level.[20]

For the *Gift* treatment, we expect that agents reciprocate the gift by the principal with increased effort. This hypothesis represents the famous gift-exchange hypothesis (Akerlof, 1982; Levine, 1998; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) which does not differentiate between different types of tasks.

H2: ***In the creative and the simple routine task, the monetary gift increases performance in periods 2 and 3 due to positive reciprocal behavior of agents towards the principal.***

In contrast, the psychological approach focusing on controlling aspects of rewards argue that rewards which are assigned independent on performance should not affect intrinsic motivation as no controlling aspect is involved in such rewards (Deci et al., 1999). A different argument has been raised, if informational aspects of rewards are important. Performance-independent rewards may decrease intrinsic motivation and thereby creative performance, because they are perceived to reward "average" or inadequate performance (Eisenberger et al., 1999). Further, they may signal that the individual who is offering the reward is uninterested in the quality of their performance (Eisenberger and Rhoades, 2001). Thus, performance-independent rewards may also crowd-out intrinsic motivation and lower creative performance in the gift treatment in period 2 and 3.

[20] Identical to period 2, a comparison of the *Tournament* to the *Feedback* treatment will allow us to disentangle the effect of mere feedback from the tournament effect. Standard economic theory would not expect any performance changes in the *Feedback* treatment, neither in period 2 nor 3. However, a number of recent economic studies show significant performance effects of relative performance feedback. Azmat and Iriberri (2010) and Blanes i Vidal and Nossol (2011), for example, find positive performance effects of feedback whereas Barankay (2011) documents detrimental effects of performance rankings. Psychological studies largely agree that the announcement of positive feedback should enhance feelings of competence, and thus intrinsic motivation (Deci et al., 1999; Vansteenkiste and Deci, 2003). Thus, recipients of positive performance feedback should reveal a positive ex-post effect on creative performance in period 3 in both, the *Tournament* as well as the *Feedback* treatment. However, a significantly lower performance in the *Tournament* than in the *Feedback* treatment in period 3 would suggest a crowding-out effect due to the financial reward. In contrast, a crowding-in effect on intrinsic motivation of the financial reward would become apparent when the *Tournament* has a positive ex-post effect on winners performance which exceeds the response of agents towards positive feedback.

4 Results

In the following, we first present descriptive statistics on the experimental results. Subsequently, we analyze treatment effects in period 2 and 3, and investigate underlying mechanisms with supplementary treatments.

4.1 Descriptive Statistics

Table 3 provides an overview of observations per treatment and tasks. Overall, our sample contains 1126 subjects, where 227 subjects were assigned the role of a principal and 899 the role of an agent.[21] As we are interested in assessing how rewards affect performance, we run sessions until the sample comprises round about 60 observations of agents, who were rewarded by their principal, for each treatment.[22] For our analysis, we mainly focus on behavior of rewarded agents, which amount to 232 in our main treatments with the slider task, 240 in our main treatments with the creative task, and 240 in four supplementary treatments which are introduced 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, see Table 3. However, as we will show in our analyses, controlling for these characteristics does not alter the results.

At the bottom of Table 3, we report 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 social preferences towards the principal. Mean performance varies between 17 and 22 in the slider task, where performance is measured as the number of correctly positioned sliders within the three minute work period. In the creative task, average performance varies between 16 and 18. It is measured as the total score for a subject's answers in the creative task (see Section 2 for details on the scoring procedure). There are no statistically significant differences between the treatments and the respective control groups in either task apart from the tournament in the slider task, which exhibits a somewhat larger performance in period 1 as compared to the control group (Wilcoxon rank-sum test, $p<0.05$).

In the following, we analyze results from our main treatments. In a simple analysis, we first look at how performance varies from period 1 to period 2 within each treatment. Figure 2 displays

[21] We excluded 9 participants in the role of agents from the analysis due to an insufficient knowledge of the German language and severe problems to understand the experiment.

[22] Principals 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 association between reward implementation and baseline performance. We will report statistics on principals' decisions on the reward implementation as well as agents behavior in case of a reward denial further below in this Section.

the change in performance from period 1 to period 2 by treatment and task. Let us first consider the slider task. In this task, all treatment groups increase their performance significantly from period 1 to period 2 (Wilcoxon signed-rank test, $p<0.01$). The control group shows a moderate increase in performance which may comprise learning effects or a response to the wage increase of 300 Taler which was exogenously given in control group, independent of a decision by the principal. The identical wage increase in the control group and the *Gift* treatment allows us to disentangle the pure effect of reciprocity towards the gift from a mere effect of the wage increase.

In order to account for the increase in performance within each treatment, we now compare the change in performance of slider-task treatments to the change in the control group. We find an increase of approximately 6 sliders in the *Gift* treatment which is significantly larger than the increase of 3 sliders in the control group (Wilcoxon rank-sum test, $p=0.07$). The *Tournament* shows the largest improvement with an average increase of more than 10 sliders (Wilcoxon rank-sum test against the control group, $p<0.01$). The *Feedback* treatment only shows a small improvement comparable and not significantly different to the change in the control group.

Now, let us consider the creative task where the pattern of results is different. Here, only agents in the *Tournament* increase their performance between periods 1 and 2. The magnitude of the increase of roughly 8 points is sizeable and highly significant in comparison to the control group (Wilcoxon rank-sum test, $p<0.01$), where performance remains unchanged. There are no notable changes in mean performance for the *Gift* and the *Feedback* treatment in the creative task, neither over periods nor in comparison to the control group.

This asymmetric pattern in the response towards the gift also carries over to period 3. Figure 3 displays performance changes from period 1 to period 3 by treatment and task. Agents in the *Gift* treatment of the slider task increase their performance significantly more than agents of the control group from period 1 to period 3 (Wilcoxon rank-sum test, $p=0.07$). In contrast, the *Gift* treatment in the creative task does not induce any performance improvement to period 3, neither compared to period 1 nor compared to the control group. Remarkably, average performance in the *Tournament* treatments remains in period 3 at a level which is still significantly higher than in period 1. However, compared to the respective control group this change is only significant in the creative task (Wilcoxon rank-sum test, $p=0.02$). The provision of feedback does not increase performance to period 3 in both treatments, neither compared to period 1 nor against the change in the respective control group.

4.2 Empirical Approach and Main Analyses

In order to make performance between the different items of the creative task as well as between the creative and the routine task directly comparable, we use standardized performance

measures for the following analysis.[23]

A frequent concern in studies using repeated observations is regression towards the mean (Barnett et al., 2005). This implies that low and high performing individuals will perform closer to the average performance in the next period irrespective of any intervention. This is also the case in our study where regression to the mean can be detected for both tasks and of similar magnitude. A participant in the control group who performs one standard deviation above the mean in period 1 performs 0.22 standard deviations worse in the slider task and 0.29 standard deviations worse in the creative task in the next period.[24] We use a regression approach, a standard technique for randomized experiments (Imbens and Wooldridge, 2009), where we address mean reversion by controlling for baseline performance. Further, this allows us to additionally to control for the some imbalances between the treatments we have seen in Table 3 and to exploit the panel structure of our experiment.

In our main specification, we regress standardized performance in period 2 as the dependent variable on treatment dummies as well as the baseline performance in period 1, using the ordinary least squares (OLS) method:

$$\begin{aligned}\text{Std. Performance period } 2_i = \beta_0 + \beta_1 \text{ (Std. Performance period } 1)_i \\ + \beta_2 \text{ Gift}_i + \beta_3 \text{ Tournament}_i + \beta_4 \text{ Feedback}_i + \epsilon_i\end{aligned} \tag{1}$$

The control group serves as the reference group. Further, we adjust standard errors for potential heteroscedasticity in all regressions.[25] 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 which does not alter our results.[26] Similarly, the exclusion of the baseline performance does not alter the interpretation of our results, as shown in Column I of Table 4. For the following analyses, we will refer to our main specification (Column II).

The *Tournament* has a substantial and significant positive effect on performance in both

[23]We use the standard approach of subtracting the mean performance of the control group (in the respective working period and task) and dividing the resulting difference by the standard deviation of the control group in the respective task. Therefore, the standardized performance of the control group has a mean of zero and a standard deviation of one. Standardization is a linear transformation, which eases the interpretation of the results, but does not affect the results per se as it does not change the shape of the distribution of the variables. Treatment dummies remain non-standardized to ease interpretation.

[24]In a simple linear regression for the control group where we regress the standardized performance in period 1 on the standardized performance in period 2, the coefficient of the standardized performance in period 1 is 0.78 in the slider task and 0.71 in the creative task (both p-values $\ll 0.01$).

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

[26]We also run more comprehensive specifications with further control variables, such as the Big Five and incentivized risk and reciprocity measures, but they do not alter our results. Column IV also includes supplementary treatments which will be discussed in Chapter 4.3.

tasks. The effect is substantially larger than that of the *Gift* or *Feedback* treatment in both tasks.[27] Moreover, the tournament effect does not differ significantly between the two types of tasks: an agent in the *Tournament* treatment increases his performances by approximately 0.7 standard deviations compared to the control group in both tasks. Therefore, we confirm a positive treatment effect of similar magnitude across tasks as predicted by standard economic theory (**H1a**). In our study, creative performance did not suffer under the tournament incentive as this would be the case if a strong crowding-out on intrinsic motivation outweighs positive incentive effects. In order to test whether the *Tournament* particularly affects highly intrinsically motivated agents, we analyzed the effect of the tournament incentive separately for agents who performed above and below average in period 1. We do not find that high and low performers respond differently to the *Tournament* in the creative task. In Section 4.4, we will further investigate whether a crowding-out effect occurs in period 3 when the incentive is withdrawn again. Moreover, the effect of the *Tournament* may vary by gender. The literature on gender effects suggests that men perform better in tournaments than women (Croson and Gneezy, 2009). Running a regression with gender-specific treatment effects (not reported) reveals that men perform slightly better in the *Tournament* than women in both tasks, but the difference is not significant.

The *Feedback* treatment increases performance by about 0.18 standard deviations which is marginally significant and does not differ between the two tasks. Hence, about one fourth of the increase that we observe in the *Tournament* treatment may be driven by concerns for relative standing while the remainder is driven by the desire to win the tournament prize, that is, the monetary incentive. Thus, although the *Tournament* effect is significantly larger than the effect in the *Feedback* treatment in both tasks, the mere announcement of a relative performance evaluation has a moderate positive effect, too.

Concerning the gift, we find diverging results between the slider and the creative task: the *Gift* treatment induces an economically and statistically significant effect only in the slider task ($p<0.05$). The coefficient for the *Gift* treatment, 0.21, is substantial but still only about 33% of the tournament effect size. This is in line with the notion that a monetary gift induces a reciprocal effort increase of agents. However, contrary to the prediction of reciprocity models, there is no evidence for reciprocity in the creative task, the coefficient is actually slightly negative. Thus, for the period 2, we reject can hypothesis **H2** that agents react reciprocal and increase their performance towards the monetary gift in both tasks. We can only confirm a positive reciprocal effect for the slider task but not for the creative task.[28] Significant differences

[27] A Z-test yields $p=0.1$ for *Gift* vs. *Tournament* in the slider task, and $p<0.05$ for all other comparisons

[28] In the experiment, we implemented opportunity costs of working via a time-out button (see Section 2.2). This provides us with the opportunity to examine whether treatment effects in the *Tournament* as well as in the *Gift* treatment are driven by a reduction in the number of breaks or whether the task performance during actual work

in the response to the *Gift* by gender are not supported by our data.

The slider task is uni-dimensional and, hence, a performance increase in this dimension is unambiguously positive for the principal. Performance in the creative task, by comparison, is multi-dimensional and a concern might be that subjects increase their score by, for example, substituting originality of their answers with quantity. We score answers in the creative task by evaluating the number of valid uses (validity) as well as their diversity (flexibility - number of categories) and originality (statistical frequency of a response). Thus, the score in the main analysis above represents a summary measure. Table 5 shows the results split up by these different dimensions. The pattern of results that we discussed for the overall score in our main analyses is reflected in all three dimensions of creativity. There is, for example, 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 of the total number of points – referred to as "originality rate" in the table – is even slightly higher in the *Tournament* treatment. That is, agents do not only increase performance by the total number of uses but also by the total number of original uses, and hence by the share of original answers of all answers. This is strong evidence against the notion that individuals substitute quantity for quality. As the originality of answers is measured by the infrequency of responses, we also checked whether our results change when we use subjective originality ratings. Therefore, we asked two research assistants blind to the treatments to evaluate the answers with regard to their creative originality. The evaluators were instructed to assign one point to answers they perceive as original and two points to answers they perceive as very original, and zero points otherwise. Using the score from this subjective originality assessment as dependent variable does not alter our results. Again, we see a positive and significant effect of the tournament and no treatment effect of the gift in period 2.[29]

It is also insightful to look at the decisions of principals regarding the reward implementation. While the majority of principals (88% in the slider task and 94% in the creative task) implemented the costless *Feedback*, about three quarters of the principals invested in the *Tournament* in both tasks. The *Gift* was implemented by exactly three quarters of the principals in the slider task. In contrast, only approximately half of the principals opted for the *Gift* in the creative task. This is suggestive evidence that a substantially higher share of principals anticipated that the *Gift* would induce a positive reciprocal reaction by their agents in the simple, routine task, but not in the creative task.[30]

---

time improved. Table B.1 in Appendix B reports results from analyses where we add the number of breaks in period 1, the change in breaks between periods 1 and 2, as well as interactions of the two to our main specification (Equation 1). The *Tournament* effect remains positive and significant even after controlling for the number of breaks (albeit much smaller). 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, an extension of actual working time.

[29] These results are available from the author upon request.

[30] We find no evidence for intra-session-correlation regarding the reward implementation. Results upon request.

We also asked principals to provide reasons for why they decided in favor of or against the implementation of the reward.[31] Roughly about half of the principal indicated that they implemented the *Tournament* in order to maximize their own payoff, in both tasks. The other half of the principals either stated efficiency concerns (maximizing the total payoff of the group) or to reward good performance with similar shares in both tasks. In case of the *Gift*, the majority of principals (approximately two thirds) indicated own profit maximization as the main reason for a positive reward decision. Thus, the expectation of positive reciprocity seems to be the main driver for the provision of the *Gift* in both tasks. However, while all remaining principals in the slider task indicated efficiency concerns, in the creative task only 14% do so, and 21% indicated that they want to treat their agents nicely. Moreover, virtually all principals who denied to provide the *Gift* in the creative task indicated that they do not expect own monetary utility form assigning the gift. Only approximately half of the principals do so in the slider task while the other half marked "other reasons".

Finally, we turn to the effort behavior of agents when the principal refused to implement the reward. These results have to interpreted with care as (1) the numbers of observations are very low for some treatments (in particular for the *Feedback* treatment), and (2) by design, the fixed payoffs for principals and agents were different when rewards were not implemented. Hence, fixed endowments are not identical across treatment and control group when looking at negative bonus decisions.[32] In total, 128 agents were affected by a negative reward decision of the principal (in our main treatments). Figure 4 shows that agents do not react negatively compared to the control group when the principal refused to implement the *Tournament*. However, we find a very strong negative effect on performance when principals do not assign the *Gift* in the slider task which is more than three times as large as the positive response. It is frequently observed that negative reciprocity is a stronger phenomenon than positive (Kube et al., 2013). But agents in the creative task only weakly reduce their performance when the *Gift* was denied. Thus, the pattern of reciprocal behavior towards the *Gift* seems to be mirrored for negative reward decisions.

The following section will further investigate the surprising asymmetry in reciprocity which we observe between the slider and the creative task.

[31] Principals could mark whether they opted in favor of the reward to (1) maximize own profits, (2) 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. In case the principal denied the reward, they could mark whether they (1) think that rewards are not profitable for their own payoff, (2) do not want agents to earn more, (3) do not want agents to earn more than themselves or (4) other reasons.

[32] In case the rewards are not implemented, both principals and agents earn a fixed wage of 300 Taler in period 2, whereas the fixed wage of principals amounts to 100 Taler and 600 Taler for agents in the control group.

4.3 Why Do We Not Observe Reciprocity Towards a Gift in the Creative Task?

In the following, we will discuss potential mechanisms that can explain why agents do react reciprocal in response to a monetary gift in the routine task, but not so in the creative task.

Various reasons may explain this asymmetry. One issue might be that subjects in the creative task use a so-called "moral wiggle room" as an excuse for not performing well. Creative tasks might be thought of as having a weaker link between effort and outcome. Hence, one might have tried hard but still not performed well. Agents who are motivated by image concerns and want to appear reciprocal (rather than be reciprocal) might therefore use this as an excuse for not increasing effort assuming that performance does not clearly reflect effort and that the principal is aware of this. However, a study by van der Weele et al. (2010) has shown that reciprocal behavior is robust to the availability of "excuses" to not perform well. Hence, this mechanism should not explain our results.

Further, it might be the case that people have higher effort costs in the creative task. This would suggest that the same reciprocal inclination would result in a smaller performance increase in the creative task. However, this explanation is incompatible with the fact that subjects increase their performance similarly (in terms of standard deviations of the control group) for the same tournament prize in both tasks. This suggests that effort costs are, in fact, very similar in both tasks.

Finally, a possible explanation might be a higher uncertainty about the payoff consequences of own effort to the principal in the creative task. That is, in the creative task it was not exactly known to agents how many Taler they generated by their effort for the principal. Although the scoring procedure was exemplified in the instructions, subjects could not perfectly observe their score, as in the slider task, because it depended on whether their ideas were judged as valid, on the concrete categories answers were assigned to as well as on the statistical frequency of their answers which was not known to subjects. It is a typical feature of creative ideas that their value is a priori not known. It needs thorough evaluation and sometimes the implementation of an idea until benefits from an idea can be observed. Prior studies have shown that the uncertainty about payoff consequences for the principal reduces reciprocity (Hennig-Schmitt et al., 2010; Englmaier and Leider, 2010).

In order to investigate whether this may explain 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 one aspect: agents additionally received the information after each period how many Taler they generated and could decide how much of this surplus they actually want to transfer to the principal.

Principals and agents were informed about this procedure in the experimental instructions. Identical to the creative *Gift* treatment, *Creative Transfer Gift* allowed the principal to assign a monetary gift of 300 Taler before the start of period 2. The treatment *Creative Transfer Control* serves as an additional control group to disentangle the reward effect from the mere effect of an announcement of the generated surplus and the transfer possibility.

Baseline performance is very similar between *Creative Transfer Control* and *Creative Transfer Gift* (approximately 21, see Table 3) but significantly higher compared to the other treatments with the creative task. This is not surprising since agents who like to perform on the task for intrinsic reasons but do not want the principal to benefit can do so and nevertheless transfer little (or nothing). In fact, only 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 amounts in between. Agent's mean transfer to the principal is higher in the *Creative Transfer Control* compared to *Creative Transfer Gift* (17.5 vs 15.1) but the difference is not statistically significant (Wilcoxon rank-sum test, p=0.19). In order to account for these initial performance difference, we control for baseline performance in all regressions.

In the *Creative Transfer* treatments, neither the *Gift* treatment nor the control group increases their performance significantly from round 1 to 2. However, if we look at transfers rather than performance, we see that agents in *Creative Transfer Gift* increase transfers significantly compared to round one (Wilcoxon signed-rank test, $p<0.01$) as well as relative to *Creative Transfer Control* (Wilcoxon rank-sum test, $p<0.05$). Figure 5 depicts the differences between all *Gift* treatments and their respective control group. The bars show coefficients from separate OLS regressions analyzing the effect of the *Gift* treatment (compared to the respective control group) on effort (transfer) in period 2 controlling for baseline performance. The figure demonstrates that agents increase their transferred effort in *Creative Transfer Gift* similarly as agents increased their effort in the *Gift* treatment of the slider task. We also analyzed the effects on transferred effort with regressions which include all other treatments. Column IV of Table 4 as well as Columns II and IV of Table adds the observations from the *Creative Transfer* treatments and show similar results for the increase of transfers in period 2.[33]

This analysis indicates that payoff uncertainty explains the discrepancy between the effect of the gift in the slider task and the creative task. The evidence contradicts arguments brought forward by Robert Eisenberger and co-authors, who suggested that performance independent rewards may crowd-out intrinsic motivation as they signal that "average" or inadequate performance is identically rewarded as outstanding performances. We do not find any evidence for a crowding-out effect of the gift, neither in period 2 nor 3. The fact that the treatment *Creative Transfer Gift* shows reciprocal behavior of similar magnitude as we observed in the slider task supports reciprocity theories as discussed in the economic literature. However, our study

[33]Reciprocal behavior in the treatment *Creative Transfer Gift* is not significantly different for males and females.

demonstrates that reciprocal behavior in creative tasks require that agents have information on the value of their ideas. In practice, this information is in most cases not available. Thus, this specific feature of creative tasks may hinder the emergence of reciprocity.

4.4 Analyses of Ex-post Effects in Period 3

In the following, we analyze how agents perform in period 3. After period 2, agents in the *Tournament* as well as in the *Feedback* treatment learned whether they did belong to the best 50% (winners) or not (losers) in period 2. Before period 3 started, agents were informed about their fixed wage of 300 Taler (which was identical to period 1), and that no further rewards or feedback will be assigned. In order to estimate treatment effects on performance in period 3, we fit the following model with OLS:

$$\begin{aligned}\text{Std. Performance period}_{it} = \beta_0 &+ \beta_1 \text{ (Std. Performance period 1)}_i + \beta_2 \text{ (Gift} \times \text{ Period 2)}_{it} \\ &+ \beta_3 \text{ (Tournament} \times \text{ Period 2)}_{it} + \beta_4 \text{ (Feedback} \times \text{ Period 2)}_{it} \\ &+ \beta_5 \text{ (Gift} \times \text{ Period 3)}_{it} + \beta_6 \text{ (Tournament} \times \text{ Period 3)}_{it} \\ &+ \beta_5 \text{ (Feedback} \times \text{ Period 3)}_{it} + \epsilon_{it}\end{aligned} \tag{2}$$

where $t \in \{2, 3\}$, treatment dummy interactions are therefore one in the respective round and zero otherwise while *Std. Performance period 1* is constant for individual i. We estimate standard errors robust to heteroscedasticity and serial correlation. The standardization eliminates any period effects (in the absence of the treatment). Therefore, we do not (need to) include dummies for periods 2 and 3. Table 6 shows the results.

For ease of interpretation, we separate the analyses by task.[34] Columns I and II show the results for the slider task, and columns III and IV depict the results for the creative task.

First, we refer to columns I and III where we report overall treatment effects in period 2 and 3. Interestingly, our main treatment effects carry over to period 3 even though there is no further announcement of a *Gift* or a *Tournament*. The reciprocal effect of the *Gift* in the slider task is significant and even slightly larger in period 3 than in period 2. Hence, reciprocity in the slider task has a persisting effect. For the creative task (column III), there is still no evidence of reciprocity in period 3. Thus, we reject hypothesis **H2** also for period 3. However, the supplementary treatment *Creative Transfer Gift* which we introduced in the previous section shows that reciprocal behavior is induced in period 2 and 3 when agents have information on the values of their ideas and can actively control the transfer of benefits from their ideas to the principal. Similar to the slider task, reciprocal behavior of agents in the treatment *Creative Transfer Gift* even increases in period 3 compared to period 2 (see Columns III and IV of Table

[34] However, our results are robust to pooling. Moreover, fixed-effects linear panel estimator or random-effects estimation show also similar results.

6).

In line with standard economic theory, performance in the *Tournament* treatment decreases in period 3 compared to period 2, as no incentive is at stake anymore. Nevertheless, performance is still larger in period 3 than in the baseline period 1 for both tasks. Learning cannot explain this finding since the performance of the control group is not significantly different between periods 1 and period 3 in any task (see Figure 3). Hence, the *Tournament* has a sustainable performance-enhancing ex-post effect. This finding speaks in favor of a crowding-in of intrinsic motivation and is inconsistent with theories on crowding-out of intrinsic motivation for the creative task. Therefore, we have to reject the hypothesis **H1b** that subjects' performance after the participation in the *Tournament* (period 3) returns to the baseline performance of period 1. The *Feedback* treatment, in comparison, does not have a lasting impact on performance. In both tasks, performance of agents in this treatment is statistically indistinguishable from the control group's performance.[35]

After period 2 ended, subjects in the *Tournament* and the *Feedback* treatment were informed whether they belonged to the best 50% of their group or not. In the *Tournament* treatment, winners were additionally informed about winning the tournament prize. As briefly discussed in Section 3, the information about own relative performance may have an effect on the performance in period 3, even in the absence of a monetary reward, due to enhanced (or lowered) feelings of competence (for instance, Deci et al., 1999) Therefore, we investigate a further specification where we look at the ex-post performance in period 3 of winners and losers of the *Tournament* and the *Feedback* treatment, separately. In Columns II and IV, we add dummy variables indicating whether subjects were informed to belong (not belong) to the best 50% of their group. The informational content about relative performance is identical in both treatments and they only differ with respect to the monetary consequence of belonging to the best 50%. Results show that positive ex-post effects in the *Tournament* are driven by winners of the *Tournament* in both tasks. The latter exhibit a significantly higher performance in period 3 than the control group. Individuals who received positive relative performance feedback in the *Feedback* treatment also show a higher performance compared to the control group. However, their increase in performance is only approximately half as big as the one of tournament winners. This indicates that positive relative performance feedback seem to enhance intrinsic motivation in both tasks whereas the effect is particularly strong when subjects additionally received a monetary prize on top of the positive feedback.[36] Recipients of negative performance feedback (in the *Feedback* as well as the *Tournament* treatment) did not change their effort in period 3 substantially compared to baseline performance. Only in the creative

[35]The coefficients of the *Tournament* and the *Feedback* treatment are marginally significantly different only in the creative task (Z-test, one-sided, $p< 0.1$).

[36]However, the coefficients for *Tournament* winners and *Feedback* winners is neither significant in the creative task nor in the slider task.

task, losers in the *Tournament* show a larger performance level in period 3 compared to their baseline.

There are several possible explanations for the positive "winner effect". A likely candidate is a heightened task-specific self-confidence which in turns increases intrinsic motivation in the respective task. This has been discussed in the psychological literature where substantial evidence show that performance-dependent rewards can crowd-in intrinsic motivation (for instance, Eisenberger et al., 1999; Eisenberger and Shanock, 2003; Vansteenkiste and Deci, 2003). Moreover, it has been argued that tangible rewards, compared to pure performance feedback, make the signal of recognizing extraordinary performance more salient. Thus, tangible rewards may be even more powerful in triggering feelings of competence, and hence, intrinsic motivation than mere positive feedback (Eisenberger et al., 1999). Alternatively, the effect could be driven by learning on the task as winners worked by definition more on the task and, hence, have gained more experience. As already mentioned, learning is unlikely to explain the effect as there is no indication in the control group that learning plays an important role for performance in these tasks. Another natural candidate to explain the positive ex-post effects is the fact that winning a tournament puts individuals in a positive mood.[37] This could affect individuals' general sense of competence and induce an heightened intrinsic motivation which is more general and should also spillover to a different task in period 3. In contrast, a crowding-in effect, as discussed in the literature, should be limited to the particular task, and hence, only affect intrinsic motivation in the same task. Since this distinction is important for the managerial implications of this study, we conducted two supplementary tournament treatments investigating spillover effects. Here, subjects worked on both tasks, the routine and the creative task, in varying order. Subjects faced one of the following task order: slider task (period 1) - creative task (period 2) - slider task (period 3), which we refer to as *SCS*, or creative task (period 1) - slider task (period 2) - creative task (period 3), which we refer to as *CSC*. Identical to the *Tournament* treatments for the creative or the slider task, agents received a fixed wage in period 1 and 3 and the principal could implement a tournament for the task in period 2. The supplementary treatments allow to study whether the positive ex-post effect of winning in the tournament only occurs in the same task or spills over to different tasks, too. Thus, we can disentangle whether observed effects are driven by changes in task-specific intrinsic motivation or stem from a general increase in intrinsic motivation which could be induced, for instance, by mood effects.

In total, 55 subjects participated in the *Tournament* treatment with the mixed tasks *Slider-Creative-Slider (SCS)*, and 46 subjects with mixed tasks *Creative-Slider-Creative (CSC)*.

[37] For a review of mood effects on performance see, for instance, Lane et al. (2005). Related to mood effects, Kräkel (2008) discuss that tournaments may induce a "joy of winning" which affects motivation.

The right columns of Table 3 shows summary statistics for these observations. There are no significant differences in baseline performance between the supplementary *Tournament* treatments and the respective control group of the main treatments.[38]

Table 7 shows the results from separate regressions for the effect of the *Tournament*, as well as of tournament winners and losers, on performance in period 3 (ex-post effect) for the slider task (Columns I and II), the creative task (Columns III and IV) and the mixed tasks, *SCS* and *CSC* (Columns V-VIII). In each regression, we include baseline performance and compare effects to the respective control group. We find no evidence for positive spillover effects compared to the control group in any of the tasks – neither an overall treatment effect nor effects for winners or losers separately. The treatment effects in the mixed tasks, *SCS* and *CSC* (Columns V-VIII), are positive but small and insignificant. That is, winning the *Tournament* does not lead to increased performance in a different task. This result supports the notion that winning a tournament leads to task-specific subsequent performance increases, for instance, via increased self-confidence and intrinsic motivation on that particular task. In contrast, more general and task-unspecific effects on intrinsic motivation, for instance via mood effects, are not supported by our data.

5 Conclusion

In today's society many routine tasks are computerized and employees increasingly face tasks that involve creative thinking. This raises the question of how to design reward schemes that foster creativity. Traditionally, economic theory on the impact of rewards abstracts from context and type of task. Hence, the large body of research on incentives for simple and routine tasks also speaks to the question of how to enhance creativity. In contrast, a comprehensive psychological literature assumes that reward effectiveness depends on the type of task, in particular on the intrinsic motivation involved in a task (Shalley et al., 2004). In this literature, opposing arguments have been discussed how intrinsic motivation involved in creative tasks is affected by monetary rewards. On the one hand, crowding-out theory (Deci, 1975; Frey, 1997) suggests that performance-dependent rewards are likely to decrease intrinsic motivation while performance independent rewards should not affect intrinsic motivation, and thus creative performance. On the other hand, a contrary approach on "crowding-in" effects (for instance, Eisenberger and Cameron, 1996; Eisenberger and Shanock, 2003) argues that performance-dependent rewards may even foster feelings of self-determination and thereby

[38]We compare the supplementary tournament treatments, *SCS* and *CSC*, to the control group from the main treatments where subjects worked on the same task in period 1 and 3. That is, we use the control group of the creative task for the treatment *CSC*, and the control group of the slider task for the treatment *SCS*. As we did not conduct additional control groups with varying tasks, we cannot control for changes in performance in period 3 which are caused by a change in the tasks per se.

increase intrinsic motivation and creative performance. In contrast, wage gifts have been reported to have detrimental effects on intrinsic motivation as they signal that average performance or inadequate performance is rewarded (Eisenberger and Rhoades, 2001). Both theories offer opposing predictions on how performance dependent or performance-independent rewards affect creative performance. Further, this raises the question whether rewards do in fact have different effects on routine and creative tasks.

This paper reports the results from a large-scale experiment that addresses these questions. We investigate the effect of both a performance-dependent tournament incentive as well as a performance-independent gift in both a creative and a routine task. We use a laboratory setting which allows us to hold constant factors that inhibit a comparison of effect sizes between tasks and rewards in the field. We observe performance over three periods, which gives us a measure of baseline performance, of performance directly after reward announcement (incentive stage in case of the tournament), and of ex-post performance. Our results suggest that a tournament prize for above average performance has a substantial positive incentive effect in both types of tasks. The effect size is similar in both the creative and the routine task. This indicates that intrinsic motivation is not crowded out in the creative task. Only about one fourth of this effect is driven by a concern for relative rank, as shown in a treatment where rank information is provided but comes without financial consequences. Thus, the monetary incentive of the tournament is largely driving the positive effects as predicted by standard economic theory. Further, the increase of creative performance by the *Tournament* can be observed in a quantitative as well as a qualitative manner, and do not come at the expense of the originality of ideas. Losers of the tournament do not show a demotivated effect in neither of the tasks. That is their performance after the tournament returns to the baseline performance level prior to the tournament announcement. In contrast, tournament winners still show a larger performance after the tournament already ended. We investigate and discuss several mechanisms which could explain this result. We conclude that this is most likely driven by a crowding-in of task-specific intrinsic motivation as suggested in the psychological literature (Eisenberger and Cameron, 1996).

Interestingly, we find an asymmetry in the effectiveness of a monetary gift between the two tasks. In line with the existing literature on reciprocity, the *Gift* successfully triggers reciprocity in the simple routine task. However, we do not observe reciprocal behavior of agents in the creative task.

A potential reason for this finding is an intransparency of generated surplus for the principal between the tasks. Agents can easily observe their score in the routine task. In contrast, the value of their creative ideas depends on their evaluation and is not perfectly observable to agents. Therefore, we conducted an additional treatment where agents receive the information on their generated surplus for the principal after each period, and are subsequently able to decide how

much of this surplus they want to transfer to the principal. We find that introducing transparency on principals' surplus in this way triggers reciprocal behavior also in the creative task. The effect size is similarly large as of the *Gift* in the routine task. It has already been demonstrated by other studies that missing information on how agents' effort translates into benefit for the principal hinders reciprocal behavior in the field (Hennig-Schmitt et al., 2010). This study provides evidence that this seems to be particularly important to consider for creative tasks where employees can generally not quantifiable the value of their ideas in an easy manner.

However, even if potential benefits from ideas are transparent, our study suggests that wage gifts are less efficient than a tournament in triggering agents creative performance. Compared to the control group, the assignment of the *Gift* did not pay-off for the principal, neither in the slider task nor in the creative task (*Creative Transfer Gift* treatment). Instituting the *Tournament*, on the other hand, was payoff-increasing in comparison to the control group in both tasks.[39] Our results imply that tournaments appear as a preferred incentive scheme to trigger creative performance as compared to wage gifts. However, strong caveats such as sabotage, high risk aversion among employees or early give-ups of low performers are issues, a firm should consider when trading these rewards types against each other.

Future work also needs to address 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, different types of creative tasks may show different effects of rewards. As we studied in particular divergent thinking as one specific type of creativity, a tournament reward might work less well for other forms of creative performance.

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

References

Abeler, J., Falk, A., Götte, L., Huffman, D., 2011. Reference points and effort provision. The American Economic Review 101 (2), 470–492.

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–Ű578.

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

Amabile, T. M., 1979. Effects of external evaluations on artistic creativity. Journal of Personality and Social Psychology 37, 221–233.

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., Iriberri, 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.

Barnett, A., Van Der Pols, J., Dobson, A., 2005. Regression to the mean: what it is and how to deal with it. International Journal of Epidemiology 34 (1), 215–220.

Bellemare, C., Shearer, B., 2009. Gift giving and worker productivity: Evidence from a firm-level experiment. Games and Economic Behavior 67 (1), 233–244.

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.

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 Litera 50 (2), 368–425.

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

Bradler, C., 2009. Social preferences under risk, ZEW Discussion Paper No. 09-077, Mannheim.

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

Brueggen, A., Strobel, M., 2007. Real effort versus chosen effort in experiments. Economics Letters 96, 232–236.

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.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic Literature, 448–474.

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., 1975. Intrinsic motivation. Plenum Press, New York.

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.

Deci, E. L., Ryan, R. M., 1980. The empirical exploration of intrinsic motivational processes. In: Berkowitz, L. (Ed.), Advances in Experimental Social Psychology. Vol. 13. Academic Press, New York, pp. 39–80.

Deci, E. L., Ryan, R. M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press, New York.

Dickinson, 1999. An experimental examination of labor supply and work intensities. Journal of Labor Economics 17 (4), 638–670.

Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268–298.

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

Ederer, F., Manso, G., 2012. Is pay-for-performance detrimental to innovation? Management Science Forthcoming.

Eisenberger, R., Cameron, J. A., 1996. Detrimental effects of reward: Reality or myth? American Psychologist 51 (11), 1153–1166.

Eisenberger, R., Pierce, W., Cameron, J., 1999. Effects of reward on intrinsic motivation–negative, neutral and positive: Comment on deci, koestner, and ryan (1999). Psychological Bulletin 125 (6), 677–691.

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

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

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

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), 1176Ű1196.

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., 1997. Not Just for the Money: An Economic Theory of Personal Motivation. Edward Elgar, Cheltenham, UK.

Frey, B. S., Jegen, R., 2001. Motivation crowding theory. Journal of Economic Surveys 15 (5), 589–611.

Gerlitz, J.-Y., Schupp, J., 2005. Zur erhebung der big-five-basierten persönlichkeitsmerkmale im soep. Tech. rep., DIW Berlin, research Notes 4.

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., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. The Journal of Economic Perspectives, 191–209.

Gough, H. G., 1979. A creative personality scale for the adjective check list. Journal of Personality and Social Psychology 37 (8), 1398.

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.

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.

Holmstrom, B., Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics, and Organization 7 (2), 24–52.

Imbens, G. W., Wooldridge, J. M., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47 (1), 5–86.

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

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.

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.

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.

Murphy, R. O., Ackermann, K. A., Handgraaf, M. J. J., 2011. Measuring social value orientation. Judgment and Decision Making 6(8), 771–781.

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.

van der Weele, J., Kulisa, J., Kosfeld, M., Friebel, 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 Emotiond 27 (4), 273–299.

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

Figures and Tables

Figure 1: Screenshot of the Slider Task

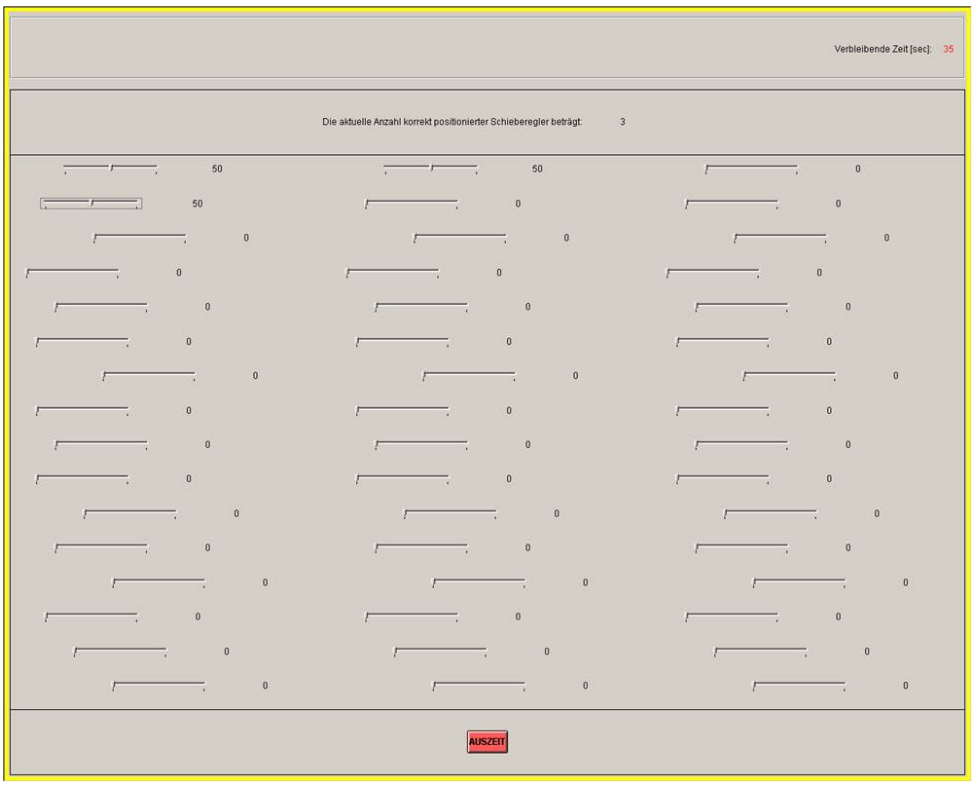


Figure 2: Difference in Performance between Period 2 and 1 by Treatment and Task

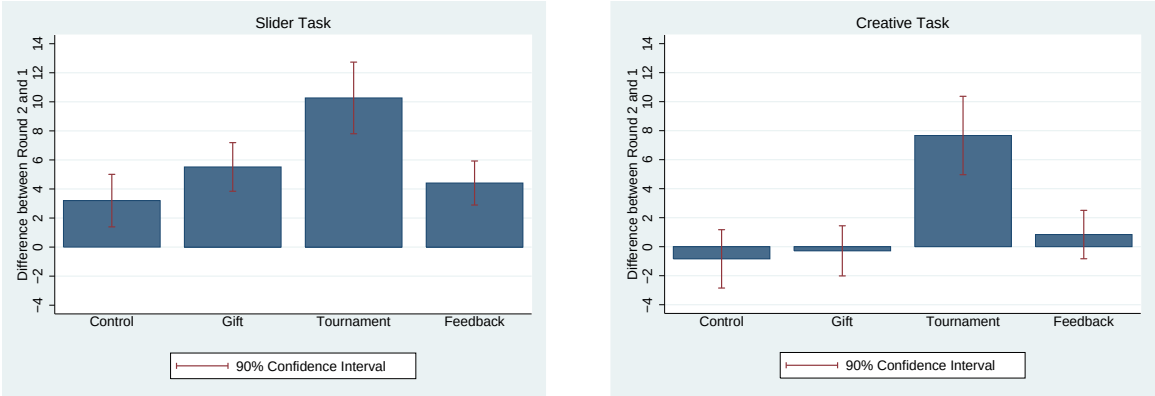
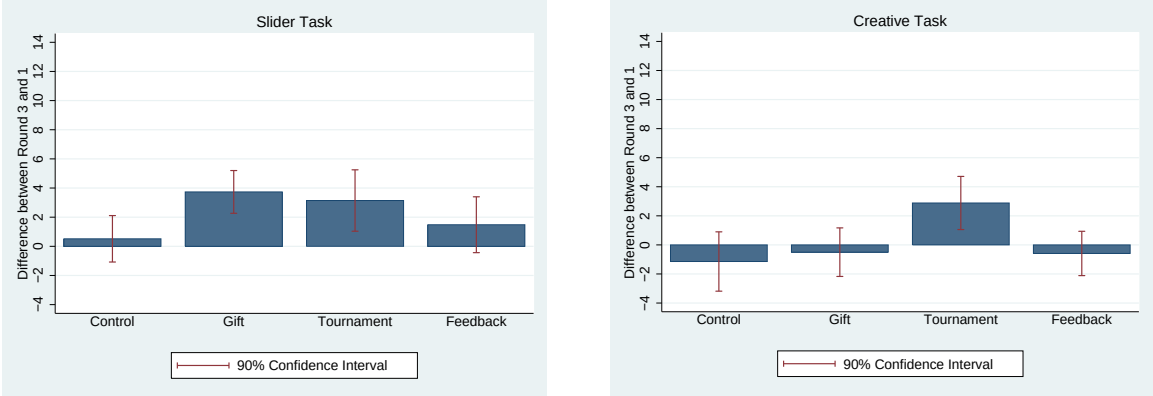
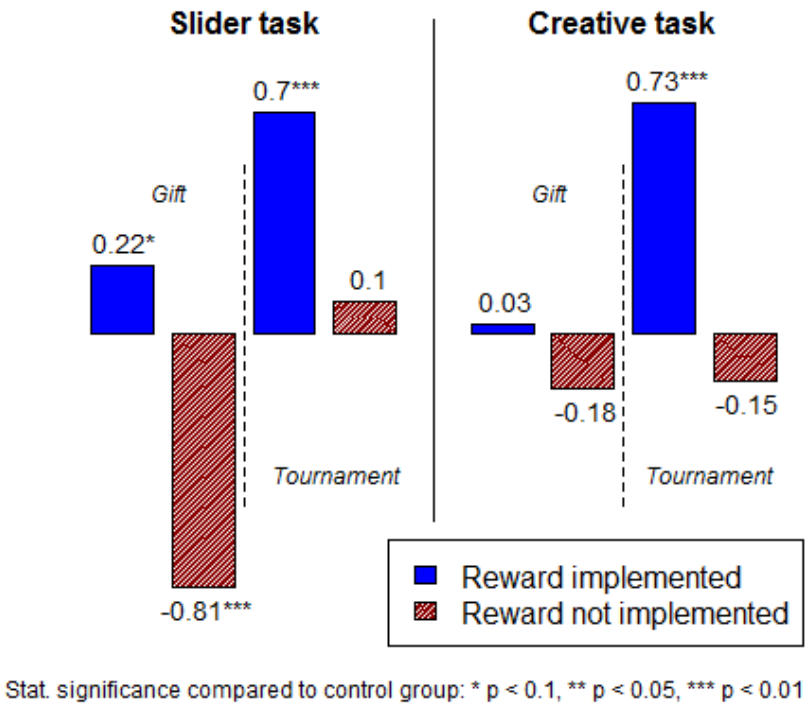


Figure 3: Difference in Performance between Period 3 and 1 by Treatment and Task



Note: The bars show the difference in mean performance between period 2 and period 1, or between period 3 and period 1, respectively. The whiskers depict 90% confidence intervals of a paired t-test.

Figure 4: Overview on Effect Sizes in Case of Reward Implementation and Refusal



Note: The bars show the estimated regression coefficients in a joint OLS regression controlling for baseline performance (for implemented rewards comparable to Table 4 Column 2). Principals are excluded from this graph.

Figure 5: Overview on Effect Sizes for All Gift Treatments

Note: The bars show the estimated regression coefficients of separate OLS regressions on standardized effort (transfer) in period 2 compared to the control group and controlling for baseline performance. The whiskers depict 90% confidence intervals of coefficients. Principals are excluded from this graph.

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

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

Table 2: Overview of Pay-offs by Treatment and Role

Fixed and Treatment-specific Payments in Taler								
	Control		Gift		Tournament		Feedback	
	Principal	Agent	Principal	Agent	Principal	Agent	Principal	Agent
Period 1 - Fixed Wage	300	300	300	300	300	300	300	300
Period 2								
Fixed Wage	100	600	300	300	300	300	100	600
Reward Costs(-)/Benefits(+)	-	-	-200	+300	-200	+0/600[a]	-	-
Total (Expected) Payoff	100	600	100	600	100	600[b]	100	600
Period 3 - Fixed Wage	300	300	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 to win, 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				
	Slider Task				Creative Task				Creative Transfer		Mixed Task (SCS)	Mixed Task (CSC)	
	C	G	T	F	C	G	T	F	C	G	T	T	**Total**
Total # of Subjects	75	100	95	80	70	135	105	90	89	96	94	97	1126
# Agents	60	80	76	64	56	108	84	72	71	76	75	77	899
# Rewarded Agents	60	60	56	56	56	56	60	68	71	68	55	46	712
													Mean
Mean Age	25	23	23*	23	24	23*	23*	23	22	23	22**	22***	23
Share of Women	58%	43%	52%	57%	52%	57%	53%	38%	56%	59%	58%	50%	53%
Economics major	50%	57%	55%	64%	59%	61%	68%	49%	66%	53%	49%	67%*	58%
Location													
Frankfurt	67%	60%	64%	64%	71%	71%	73%	59%	56%	54%	27%***	26%***	58%
Mannheim	33%	40%	36%	36%	29%	29%	27%	41%	44%	46%	22%	22%	34%
Heidelberg	0	0	0	0	0	0	0	0	0	0	51%***	52%***	7%
Baseline Performance	16.6	18.9	22.0**	20.3	16.8	16.0	17.8	17.9	20.7	21.0	19.5	17.9	18.8
	(12.5)	(14.1)	(12.8)	(11.4)	(11.7)	(10.7)	(10.9)	(10.4)	(12.0)	(10.9)	(13.1)	(10.4)	(11.8)

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, hence, the number of subjects that we use in the statistical analysis. All stated means and percentages refer to the number of rewarded agents. 32 Observations were dropped because of problems wthin the sessions (SCS = 13; CSC = 2; Creative Transfer Control = 1; Creative Transfer Gift = 16).
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 two-sample Wilcoxon-Mann-Whitney 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	-0.021	0.023	-0.002	0.017
	(0.144)	(0.099)	(0.102)	(0.100)
Tournament Treatment	0.791***	0.734***	0.731***	0.727***
	(0.173)	(0.142)	(0.142)	(0.142)
Feedback Treatment	0.235*	0.172*	0.181*	0.177*
	(0.131)	(0.094)	(0.106)	(0.103)
Gift x Slider-Task	0.377**	0.200*	0.197*	0.180*
	(0.170)	(0.107)	(0.108)	(0.108)
Tournament x Slider-Task	0.170	-0.085	-0.059	-0.060
	(0.178)	(0.166)	(0.167)	(0.167)
Feedback x Slider-Task	0.145	-0.007	-0.009	0.008
	(0.169)	(0.107)	(0.114)	(0.112)
Creative Transfer Treatment				0.083
				(0.105)
Creative Transfer x Gift				0.249**
				(0.115)
Baseline		0.632***	0.629***	0.654***
		(0.063)	(0.062)	(0.046)
Baseline x Slider-Task		0.095	0.102	0.072
		(0.079)	(0.080)	(0.067)
Intercept	0.000	0.000	0.581	0.180
	(0.093)	(0.062)	(0.496)	(0.451)
Controls	NO	NO	YES	YES
Observations	472	472	472	611
R^2	0.117	0.554	0.563	0.553

Note: This table reports the estimated OLS coefficients from Equation 1. The dependent variable is the standardized performance in period 2 for both tasks. Heteroscedastic-robust standard errors including a degree of freedom correction are reported in parentheses. Additional control variables are age, age squared, sex, location, field of study as well as for specific time effects (semester period, semester break, exam period). Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates contain data from agents in groups where the principal implemented the reward.

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

	Creative Task				
	Aggregated	*Validity*	*Flexibility*	*Originality*	*Originality Rate*[1]
	I	II	III	IV	V
Gift	0.028	0.049	0.055	-0.058	-0.015
	(0.122)	(0.126)	(0.116)	(0.146)	(0.023)
Tournament	0.734***	0.781***	0.590***	0.725***	0.046**
	(0.160)	(0.168)	(0.144)	(0.182)	(0.020)
Feedback	0.172	0.151	0.162	0.177	0.015
	(0.118)	(0.121)	(0.113)	(0.147)	(0.021)
Period 1	0.632***	0.615***	0.620***	0.537***	0.186**
	(0.063)	(0.064)	(0.050)	(0.077)	(0.072)
Intercept	0.000	0.000	0.000	0.000	0.154***
	(0.094)	(-0.099)	(0.092)	(0.109)	(0.023)
Observations	240	240	240	240	215
R^2	0.463	0.44	0.498	0.267	0.072

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. In Column V, we show the effect of the treatments on the originality rate, that is, the percentage of points earned for originality of the overall creativity points of an individual. Heteroscedastic-robust standard errors are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates contain data from agents in groups where the principal implemented the reward.

[1] The originality rate equals achieved originality points divided by the total number of points for a subject's answers. The sample size is reduced to 215 observations because we cannot calculate the originality rate (or it's baseline) for individuals who do not provide any answers in either period 1 or period 2.

Table 6: Ex-post Treatment and Spillover Effects

	Slider		**Creative**	
	I	II	III	IV
Period 3: Gift	0.262**	0.267***	0.028	0.028
	(0.101)	(0.102)	(0.104)	(0.104)
Period 3: Tournament	0.307**		0.356***	
	(0.139)		(0.116)	
Period 3: Feedback	0.132		0.128	
	(0.125)		(0.118)	
Period 3: Creative Transfer			-0.028	-0.027
			(0.115)	(0.115)
Period 3: Creative Transfer x Gift			0.358***	0.356***
			(0.108)	(0.108)
Period 3: Tournament Winner		0.600***		0.474***
		(0.172)		(0.148)
Period 3: Tournament Loser		0.025		0.240*
		(0.152)		(0.132)
Period 3: Positive Relative Feedback		0.350**		0.263*
		(0.175)		(0.136)
Period 3: Negative Relative Feedback		-0.063		-0.002
		(0.133)		(0.133)
Period 2: Gift	0.187*	0.192*	0.020	0.020
	(0.103)	(0.102)	(0.110)	(0.110)
Period 2: Tournament	0.642***	0.659***	0.744***	0.746***
	(0.137)	(0.135)	(0.151)	(0.151)
Period 2: Feedback	0.146	0.158	0.227*	0.229**
	(0.108)	(0.110)	(0.116)	(0.116)
Period 2: Creative Transfer			0.081	0.083
			(0.116)	(0.117)
Period 2: Creative Transfer x Gift			0.285**	0.284**
			(0.118)	(0.118)
Controls	YES		YES	
Baseline	YES		YES	
Intercept	YES		YES	
N x T	464	464	758	758
R^2	0.672	0.684	0.508	0.511

Note: This table reports OLS estimates of standardized performances in periods 2 and 3 according to Equation 2. Performance is measured as the number of correctly positioned sliders and the score achieved in the unusual uses, creativity task. Cluster-robust standard errors (by participant) are reported in parentheses. Additional control variables are age, age squared, sex, location, field of study as well as specific time periods (semester period, semester break, exam period). Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates contain data from agents in groups where the principal implemented the reward.

Table 7: Ex-post Treatment Effects of the Tournament in Period 3 by Task

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

Note: This table reports OLS estimates of standardized performances in periods 3. Performance is measured as the number of correctly positioned sliders and the score in the unusual uses task for the creative task, respectively. Heteroscedastic-robust standard errors including a degree of freedom correction are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates contain data from agents in groups where the principal implemented the reward.

Appendix A - Instructions (Translated from German)

General Instructions

Please read the instructions carefully.[40] 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 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 your decisions and the decisions of other participants. All Taler that you earn in the experiment will be exchanged into 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. You act completely anonymously.

Assignment of roles

At the beginning of the experiment a random draw by the computer will decide whether you are assigned the role 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 also remain the same for the entire experiment.

Structure of the experiment

The experiment consists of two parts. Part 1 is the actual experiment. Employees will only interact with the employer in Part 1. In Part 2, only employees will go through further decision tasks. The instructions for this part of the experiment will later be shown on the computer screens. Subsequent to Part 1 and 2 of the experiment, you will be asked to fill out a questionnaire, which both employers and employees will receive. Please find the instructions for Part 1 below.

[40] In order to improve readability, we refrain from the simultaneous use of the feminine and the masculine form of speech. All denotations which refer to persons apply to both sexes.

Instructions Part 1

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 rounds start.

Variable payment

Employers will receive an additional variable payment, which is calculated from the work performance of the four employees of his group. To that end, the employees will be asked to complete a simple task in every round. The task will be the same in every round. The employer's payment is calculated from the total work performance of all employees in all three tasks. Until the end of the experiment, the employer will not receive any information on how much his four employees of his group produced over the three work rounds and how much he will earn. The employees are free to decide how much effort they exert, and thus raise their employer's payment. The employees' payment will not be influenced by their work performance. Additionally, employees have the option to press a time-out button during their working time. This button will lock the screen for 20 seconds so no entries can be made during that time. The employee receives 5 Taler for each time-out. When less than 20 seconds remain, the button can no longer be pressed. The employer will be shown the same tasks in all three rounds as well. The employer's work performance, however, will not influence his payment or the employees' payment. The employer is able to press the time-out button, although activation has no influence on his payment. The employer cannot earn additional Taler during time-out.

[Task description on creative or simple routine task]

The employer will not receive any information on the amount of Taler each employee generated. Only the total payment will be announced to the employer at the end of the experiment. The employees will not be informed on the total amount of Taler which was generated by themselves or other participants, either. All participants will be called individually for payment. Please remain at 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.]

Task description [Creative task]:

As described above, the first part of the experiment consists of three rounds. In each round, participants receive a task which 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 the tire. You can also list uses which 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 task. There will be a break of a few minutes after every round.

Task evaluation

The employer receives 5 Taler per valid answer. Answers are "valid" when they are practicable and 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. An answer is considered "original", if only few people think of it. For very original (very rare) answers, the employer receives 10 extra Taler. To this end, the answers are compared to a catalog of answers, which has been prepared using the answers of more than 100 test persons. Furthermore, the answers can be assigned to different 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 a wheel" and results in 5 extra Taler. The answer "swing seat" is a different category (category "toys") and yields 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 across the three rounds.

Task description [Simple routine task]:

Every round's task will be to move so-called sliders on the computer screen with the mouse. 48 sliders will be shown on each computer screen, which are to be moved to position 50. Each slider starts at position 0 and can be moved arbitrarily up to position 100. A number shown to the right of each slider indicates the current slider position. You can adjust the slider position as often as you like. Your working time is set to 180 seconds per round. If you have managed to move all sliders to the correct position, you can access a further screen, which shows another 48 sliders. Thus, you can move a maximum of 96 sliders in 180 seconds to the correct position. 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 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. The test round does not have any influence on your or the employer's payment. For each slider which is correctly positioned in the three rounds, the employer received 5 Taler. At the end of the experiment, the 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.

Appendix B - Supplementary Tables

Table B.1: Treatment Effects in Period 2 – Controlling for Working Breaks

	Slider Task I	Creative Task II
Gift Treatment	0.001	-0.062
	(0.057)	(0.109)
Tournament Treatment	0.123**	0.251**
	(0.062)	(0.123)
Feedback Treatment	0.086	-0.017
	(0.065)	(0.101)
1 to 3 Breaks (Period 1)	-0.181*	-0.237*
	(0.100)	(0.139)
4 to 7 Breaks (Period 1)	-0.461***	-0.211
	(0.157)	(0.143)
8 or more Breaks (Period 1)	-0.699***	-0.679***
	(0.207)	(0.172)
Difference between the	-0.573***	-0.218***
Breaks in Periods 2 and 1	(0.192)	(0.057)
1 to 3 Breaks (Period 1)	0.312	-0.031
× Break Difference	(0.196)	(0.068)
4 to 7 Breaks (Period 1)	0.289	0.030
× Break Difference	(0.194)	(0.068)
8 or more Breaks (Period 1)	0.253	-0.065
× Break Difference	(0.193)	(0.076)
Standardized	0.759***	0.666***
Performance in Period 1	(0.079)	(0.064)
Intercept	0.197*	0.343**
	(0.110)	(0.148)
Observations	232	240
R^2	0.904	0.664

Note: This table reports OLS coefficient estimates of the standardized performance in period 2. Performance is measured by the number of sliders which are correctly positioned or the score achieved for unusual uses in the creative task, respectively. Heteroscedasticity-robust standard errors including a degree of freedom correction are reported in parentheses. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All estimates contain data from agents in groups where the principal implemented the reward.