

Discussion Paper No. 15-021

**How Creative Are You?  
– An Experimental Study on  
Self-Selection in a Competitive Incentive  
Scheme for Creative Performance**

Christiane Bradler

**ZEW**

Zentrum für Europäische  
Wirtschaftsforschung GmbH

Centre for European  
Economic Research

Discussion Paper No. 15-021

**How Creative Are You?  
– An Experimental Study on  
Self-Selection in a Competitive Incentive  
Scheme for Creative Performance**

Christiane Bradler

Download this ZEW Discussion Paper from our ftp server:  
<http://ftp.zew.de/pub/zew-docs/dp/dp15021.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

---

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

# How Creative Are You? - An Experimental Study on Self-Selection in a Competitive Incentive Scheme for Creative Performance \*

Christiane Bradler<sup>†</sup>

March 2015

## Abstract

Economic theory suggests that performance pay may serve as an effective screening device to attract productive agents. The existing evidence on the self-selection of agents is largely limited to job tasks where performance is driven by routine, well-defined procedures. This study presents evidence for a creative task and studies how agents self-select into a tournament-based scheme vs. a fixed pay scheme. The experiment allows for the measurement of creative productivity, risk preferences, self-assessments, gender, and other socio-economic characteristics such as the Big Five personality traits. Results show that the two payment schemes systematically attract agents with different characteristics. However, results differ fundamentally from previously found patterns. Agents did not self-select into the tournament scheme according to their creative productivity, but only according to their risk attitudes and self-assessments. The reason for the absence of a selection of the most creative agents into the tournament is that there exist substantial misjudgments of relative creative productivity. Further evidence from a representative German survey data set provides additional support for the experimental results suggesting external validity.

JEL Classification: C91, D03, J33, M52

Key words: performance pay, tournaments, selection, sorting, creativity, experiment

---

\*Acknowledgments: I would like to gratefully acknowledge financial support by the framework program “Economics of Science” by the Federal Ministry of Education and Research, Germany. Further, I thank Arjan Non, Daniel Herbold, Michael Kosfeld, Steve Levitt, Dirk Sliwka, Nick Zubanov and further conference participants for valuable comments and discussions. Julia Teufel, Dominic Heinz as well as Vera Schmitz provided excellent research assistance.

<sup>†</sup>ZEW Centre for European Economic Research, Mannheim. E-mail: bradler@zew.de.

# 1 Introduction

As organizations usually cannot perfectly observe the ability of agents during the hiring process, principal agent theory suggests that performance-based compensation schemes can be used as a screening device to attract high ability types (Stiglitz, 1975; Prendergast, 1999). Assume a performance-based pay scheme which offers high ability types the opportunity to earn more while low ability types are subject to earning less than under an alternative fixed pay scheme. If agents can self-assess their ability accurately, they will self-select into the pay scheme which maximizes their expected utility, i.e., high ability types would prefer the performance-based pay and low ability types the fixed pay contract. Several empirical studies confirm that performance pay, for instance in the form of tournament bonuses or piece rate contracts, induces a substantial self-selection of relatively productive agents (see, for instance, Lazear, 2000; Banker et al., 2000; Dohmen and Falk, 2011). However, the existing evidence on self-selection into pay for performance is largely limited to tasks which are well-defined routine tasks such as installing windshields in cars (Lazear, 2000), selling products (Banker et al., 2000), or solving simple multiplication problems (Dohmen and Falk, 2011).

This paper investigates whether the finding that more productive agents sort into a tournament-based pay scheme (vs. a fixed pay scheme), as shown in Dohmen and Falk (2011), generalizes to a creative task. This research question is important because (1) the share of jobs which require creative thinking is strongly increasing (Florida, 2004), and (2) creative tasks may fundamentally differ from routine tasks with respect to an individual's ability to self-assess their own (relative) productivity (Dunning et al., 2004), which is an important requirement for an individual's self-selection in a compensation scheme. The present study observes people's self-assessments on (relative) performance in a creative task, and investigates how these affect people's sorting in a pay for performance scheme.

Moreover, pay for performance has been argued to be detrimental to creative performance by crowding-out intrinsic motivation (for an overview see, for instance, Amabile, 1996a; Deci et al., 1999).<sup>1</sup> If highly intrinsically motivated agents anticipate such an effect, they should be less likely to self-select into a pay for performance scheme (vs. a fixed wage alternative). The study can shed light on the question of whether highly intrinsically motivated agents are more or less likely to sort into pay for performance (vs. a fixed wage) in a creative task.

---

<sup>1</sup>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 people's perception about why they perform a task: own behavior is attributed to the reward and not to the enjoyment of the activity itself (Deci, 1972; Lepper et al., 1973). Another prominent approach posits that the implementation of a reward for creative tasks can be perceived as controlling, reducing perceived autonomy and feelings of competence. This, in turn, lowers people's intrinsic task interest and finally creative output (Amabile, 1983; Deci and Ryan, 1985; Hennessey and Amabile, 1988).

During the last decades, technological change gave rise to a considerable shift from routine tasks to non-routine tasks where an important share of the latter demand problem-solving capability and creativity.<sup>2</sup> In the US, 20 million new jobs were added between 1985 and 2005 in the creative economy which goes beyond sectors such as arts, design, culture or entertainment as it also refers to all research-oriented and knowledge intense professions (Florida, 2006). As globalization and digitization of markets accelerates the competition for innovations, organizations rely on creative talents (Company, 2013) who are able to generate valuable ideas as important drivers for innovations.<sup>3</sup>

One possibility for firms to attract creative talents is to offer incentives for high creative performance. In fact, performance pay is similarly widespread in the creative as in the non-creative sector.<sup>4</sup> However, in the creative sector performance is typically difficult to measure and depends on subjective evaluations. That is why creative performance is often aligned to relative performance. For instance, if firms face the question of which design to choose, which marketing strategy to follow or which new product to develop, it is much easier to use a tournament-based bonus to reward the employee with the best creative output rather than to evaluate creative output in absolute terms.<sup>5</sup>

An important requirement for individuals' self-selection into a tournament-based pay scheme is their ability to adequately self-assess their relative productivity. Dohmen and Falk (2011) observe that agents have a good knowledge of their relative productivity in a simple math task. Here, agents were sorted into a tournament-based pay scheme (vs. fixed pay scheme) efficiently, in the sense that the more productive agents chose the tournament while the less productive agents self-selected into the fixed pay. Several reasons exist to explain why self-assessments are likely to be accurate in routine tasks such as finding the solution of simple math equations. Routine tasks can be solved according to clearly-defined procedures. Quantity is typically easy to observe and quality of output is well measurable as unique solutions exist. The availability of these objective performance measures allows individuals to observe their relative performance

---

<sup>2</sup>Routine tasks typically follow clearly defined procedures and can also be executed by a machine or computer. Non-routine manual tasks demand workers' physical capabilities, visual recognition, and communication while non-routine cognitive tasks may require problem-solving or creativity. The latter are typical in professional, managerial, technical and creative occupations and cannot be accomplished by computers (Autor et al., 2003; Acemoglu and Autor, 2011).

<sup>3</sup>"[...], creativity is simply the production of novel, appropriate ideas in any realm of human activity, from science, to arts, to education, to business, to everyday life. [...] Creativity is the first step in innovation, which is the successful implementation of those novel, appropriate ideas." (Amabile, 1997, p. 40).

<sup>4</sup>Lemieux et al. (2009), for instance, shows that performance-based pay schemes are generally observed across all industry sectors and for all occupational types. Own analyses with a large German data set indicate that about a quarter of all employees in the creative as well as in the non-creative sector earn a monthly wage or yearly bonus that is performance-based. See Section 5 for details.

<sup>5</sup>Lazear and Rosen (1981) already argued that it is less costly to rank employees if absolute performance is difficult to measure. Even if it is possible to align bonuses to sales, tournaments might be favorable to absolute performance evaluations to assure employees against common shocks on the market.

in a transparent way. In contrast, an established literature in psychological science reports that self-assessments are much more likely to reveal substantial biases in tasks which are more complex or non-routine (see, e.g., Dunning et al., 2004 and references therein).<sup>6</sup> Within the range of more complex or non-routine tasks, creative tasks have specific features which make them particularly prone to misjudgments in relative self-assessments. First, no well-defined procedure exists how to generate an extraordinary idea. As a consequence, it can hardly be trained to be creative. For instance, researchers may acquire theoretical knowledge as well as practical abilities during their educational career which help them to conduct research, but no guideline exists which they can follow in order to identify a great research idea or to make a ground-breaking invention. Second, objective standards are often missing and information on alternative solutions are not available. Therefore, evaluations remain highly subjective and feedback is more ambiguous. This can cause an “error of omission” (see Dunning et al., 2004, and references therein). That is, people tend to ignore available information from feedback and adhere to their own self-perception. The question of whether people rather overestimate or underestimate their own performance has been found to depend on the perception of task difficulty. People view themselves as above average performer in domains which are viewed as easy (although they are complex), such as driving a car; and in contrast, the bias in self-assessments reverses if people view the domain as difficult (Kruger, 1999; Moore and Small, 2007). In a creative task, it remains an open question as to whether people rather tend to under- or to overestimate their relative ability in a way which hinders pay for performance schemes to attract the most creative talents. The present study will shed light on this question.

This paper presents results from a laboratory experiment and complementary evidence from a German representative field data set on agents’ self-selection into performance pay. A laboratory experiment offers the possibility to study the causal impact of creative performance, self-assessments, risk attitudes, and other individual characteristics on an agent’s sorting decision into a tournament scheme vs. a fixed-pay scheme. The structure of the experimental design is aligned to the study by Dohmen and Falk (2011) to directly compare results to a routine task. In order to observe individuals’ creativity, subjects were asked to perform a task called “unusual uses” which is known as an established creativity test. Specifically, it measures divergent thinking and has been demonstrated to consistently predict who will produce novel and useful ideas (Batey and Furnham, 2006; Guilford, 1959, 1967). In the unusual uses task,

---

<sup>6</sup>Most evidence on self-assessments in the work context did not explicitly focus on creative tasks but on complex tasks such as in managerial and professional jobs which comprise also creative jobs. Evidence by Stajkovic and Luthans (1998), for instance, shows that the correlation between how people expect to perform and how they actually perform in the workplace is weak in tasks which can be defined as rather complex, that is, tasks which require a combination of different skills, such as knowledge, cognitive ability, memory capacity, behavioral facility, information processing, persistence, and physical effort (Bandura, 1986). Similarly, a meta-analysis by Harris and Schaubroeck (1988) found that the relationship between self-ratings of employees’ and ratings of their supervisors were particularly low when tasks were less well-defined managerial or professional jobs than blue-collar or service jobs.

subjects are requested to think of as many, as different and as original uses of an ordinary object, such as a tin can. The creativity score is composed of largely objective evaluation criteria, such as the number of responses, the variety in responses and the relative frequency of a response as a measure for originality. After the creative output was measured in the first part of the experiment, subjects were asked to choose their preferred reward scheme for the next unusual uses task (with a different unknown object). They could either opt for a tournament where they could win a large tournament prize or nothing, or they could choose a fixed payment which was substantially smaller than the tournament prize. During the experiment, self-assessments as well as self-reports on perceived fun, effort, stress, and exhaustion were elicited. Further, risk preferences and other personality measures, such as the Big Five personality traits, were measured at the end of the experiment. Thus, similar to Dohmen and Falk (2011), the design allows to study multidimensional sorting with respect to a comprehensive set of personal characteristics.

Results show that the two different compensation schemes systematically attract different types of agents. In line with previous evidence, the results reveal that the more risk tolerant agents are and the more optimistically they self-assess their relative ability, the more likely they are to opt for the tournament. However, in contrast to previously observed sorting patterns in routine tasks, I do not find that the more productive agents in the creative task self-selected into the tournament. It can be shown that substantial misjudgments in agents self-assessments can explain why the pay for performance scheme does not attract the high ability types. Many agents hold wrong beliefs about their relative ability in the creative task. The specific features of creative tasks are likely candidates to explain these substantial flaws in self-assessments. There is no evidence of a crowding-out effect of intrinsic motivation that could explain the results. The opposite is true in the sense that agents who indicate a relatively high level of enjoyment during the task are more likely to opt for the tournament. However, enjoyment during the task as an indicator for intrinsic motivation is only weakly related to creative performance. Moreover, it can be shown that the absence of sorting according to creative ability cannot be explained by the fact that agents may perceive creative output as more risky.

Finally, I present complementary evidence from a German field data set which supports the results from the experiment. Here, I use information on the Big Five personality trait openness which is an established predictor for creativity (for instance, McCrae, 1987). Results show that people with a relatively high score in openness are more likely to work in the creative sector, however, they are not more likely to be found in a job with performance pay within the creative sector.

The findings of the study have important implications for HR managers and inform decision-makers which types of workers they may or may not attract with a tournament-based compensation scheme in jobs that rely heavily on creative performance.

The paper is structured as follows: the next section gives an overview on the related literature. Section 3 introduces the creative task and the experimental design. Section 4 presents the results of the experiment and Section 5 provides additional evidence from an analysis with field data to check the external validity of the experimental findings. Finally, Section 6 concludes.

## 2 Related Literature

A myriad of studies deal with the effects of performance pay in general, and tournaments in particular. However, these studies predominantly investigated the incentive effects on work performance (for an overview see, e.g., Bonner et al., 2000; Gneezy et al., 2011). So far, only a limited number of empirical studies looked at the effects of self-selection into performance pay schemes. Most of the previous evidence on sorting in performance pay schemes is based on absolute performance measures such as piece rate contracts. For instance, Lazear (2000) reports large productivity gains in a firm which replaced fixed wages with a linear piece rate scheme. He finds that a substantial share of the productivity gain goes beyond an incentive effect and can be attributed to the fact that relatively productive agents systematically selected themselves in the firm while less productive agents were more likely to leave the firm. Similar sorting patterns of the more able workers into bonus contracts with absolute performance goals have been shown by Banker et al. (2000). Further, Eriksson and Villeval (2008) and Cadsby et al. (2007) provide experimental evidence demonstrating that relatively productive agents are more likely to choose a piece rate vs. a fixed wage scheme compared to less productive agents.

More recently, Dohmen and Falk (2011) extended the evidence by conducting an experiment where they study sorting in performance pay for piece rates, revenue sharing contracts, and tournaments. As in previous studies, they confirmed that the most productive workers are significantly more likely to sort into each of these variable pay schemes while less productive workers were more likely to choose a fixed pay alternative. Dohmen and Falk (2011) further investigate multidimensional sorting and study how other factors such as agents' risk preferences, self-confidence, and personal characteristics affect the sorting decision. They find that agents who are relatively risk tolerant and self-confident with respect to their relative performance were more likely to self-select into performance pay. Economic theory predicts that risk-aversion is a constraining factor for the provision of incentives in firms (see Prendergast, 2000 and references therein). Empirical evidence supports theory and demonstrates that variable pay schemes attract significantly more risk tolerant agents (e.g., see Eriksson et al., 2009; Grund and Sliwka, 2010; Bellemare and Shearer, 2010; Cornelissen et al., 2011). In addition, it has also been shown that performance-contingent pay schemes are typically observed with higher average pay levels to compensate workers for the risk induced by the larger variability in earnings (Grund and Sliwka, 2010). For organizations, the selection

of relatively risk tolerant agents can be beneficial, for instance, when investments in new risky strategies or technologies are desirable.

A comprehensive literature on gender differences in preferences reports that women are more risk and competitiveness averse (Croson and Gneezy, 2009) which has been related to the fact that women are less likely to sort into variable pay schemes. In the study by Dohmen and Falk (2011), women preferred significantly more often a fixed payment over a variable pay scheme compared to men, which was mainly explained by their risk aversion. Niederle and Vesterlund (2007) let subjects choose between a piece rate and a tournament and find that women shy away from competition by opting for the piece rate significantly more often than men.

Moreover, social preferences can be a relevant sorting dimension for firms to increase cooperation within the organization and to avoid free-riding in teams (Fehr and Fischbacher, 2002; Kosfeld and von Siemens, 2011). However, Dohmen and Falk (2011) found that neither trust nor reciprocity determined the sorting decision into a tournament vs. a fixed pay contract in their study.

Finally, an experiment by Camerer and Lovallo (1999) shows that sorting into a tournament with performance pay is significantly affected by overconfidence of subjects.<sup>7</sup> In a knowledge quiz task, subjects overestimated their relative ability and entered a tournament more frequently compared to a condition where payoffs depend on a chance device. Based on this finding, they argue that agents who compete based on skills are insufficiently sensitive to the quality of competition and reveal a “reference group neglect”.

Overall, the evidence suggests that agents’ self-selection with respect to productivity and other individual characteristics fundamentally affects the composition of a firm’s workforce.

### 3 Experimental Approach and Design

Several reasons exist why a laboratory experiment is an ideal way to study agents’ sorting into particular compensation schemes. In practice, important sorting factors such as individuals’ ability, risk attitudes, self-perception of own (relative) productivity or other personal characteristics are rarely observable. Even if information on these factors is available,

---

<sup>7</sup>More evidence for the effect of overconfidence on contract selection exists for non-competitive piece rate and bonus contracts. For instance, in Larkin and Leider (2012) subjects performed in simple multiplication problems and trivia questions. They find that substantial share of subjects is overconfident and self-select into a convex piece rate scheme although the alternative of a linear piece rate scheme would have been payoff maximizing for them. Hyatt and Taylor (2008) show that incomplete knowledge about personal capabilities lead subjects to overestimate their absolute performance. In their study, many subjects opted for a contract which assigns a bonus for exceeding a fixed performance standard although their payoff would have been maximized by a fixed pay alternative. Only when participants learned about their actual performance capabilities by receiving absolute feedback before choosing the contract, low performers favored the fixed pay contract more often and increased their earnings.

it is very difficult to identify the causal impact of these factors on agents' self-selection. The implementation of specific HR practices, such as a new compensation scheme, is not exogenous and confounding factors may mislead causal inferences. Moreover, it will take some time until the endogenous composition of the workforce changes due to the implementation of a new compensation scheme. A lot of things may happen in between which simultaneously affect who is going to enter and who is going to leave the organization. This time delay renders infeasible a clear-cut distinction between the mere effect of the change in the compensation scheme on the workforce composition and the impact of other factors.

### 3.1 The Creative Task

In the experiment, subjects were asked to perform on a “real effort” task which requires creativity. The task is called “unusual uses task” and adopted from established creativity tests. Originally introduced by the Guilford’s Alternative Uses Task (Guilford, 1967), it was later on also integrated in the *Torrance Test of Creative Thinking* (Torrance, 1968, 1998), which is known as 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 (Sternberg, 1999; Kim, 2006).<sup>8</sup> An example of the task is as follows: “Please list as many, as different and as unusual uses of a tin can as you can think of! Do not restrict yourself to a specific size or shape. You can also think of uses which require several tin cans. Do not limit yourself to uses which you know but think of as many unusual uses as possible!”.

This task captures an essential aspect of many research-oriented and knowledge-intense professions: recombining existing bits of knowledge in novel ways (Burt, 2004; Simonton, 2004). It requires divergent thinking, also known as the ability “to think out-of-the-box”, what is seen as one of the most important components of the creative process (Runco, 1991). It is suitable to be implemented in the laboratory as it does not require previous knowledge and both the task and the scoring procedure are easy to understand.

At the beginning of the experiment, subjects received detailed instructions about the task and how it would be scored. Before the experiment started, subjects could test the task for one minute. In this test period, they were asked to think of unusual uses for the item “old tire” to familiarize themselves with the task.

During the experiment, subjects worked on the three items (paper, cord, tin can) in separate steps of the experiment. Subjects are allowed to work for three minutes on each item. Previous evidence by Bradler et al. (2014) shows that subjects who worked on the identical items

---

<sup>8</sup>While the *Torrance Test of Creative Thinking* aims to assess the overall creative potential of a person and comprises several figural and verbal elements, the unusual uses task captures the type of creativity which is the most relevant for the non-routine, knowledge intense jobs of the creative class.

sequentially with a working time of three minutes yield scores that are highly and significantly correlated across the items without an indication of significant learning or fatigue effects.<sup>9</sup>

Research assistants evaluated responses immediately after the working time of 3 minutes and subjects received feedback according to their achieved total points. Subjects' responses to the unusual uses task are evaluated according to three criteria: validity, flexibility, and originality (Guilford, 1959; Torrance, 1968).<sup>10</sup> Validity refers to the number of valid answers of "unusual uses" which are rather leniently defined as any relevant use other than a usual use. In the case of a tin can, the usual use is defined as a container for preserving food or other products which are typically sold in tin cans such as paint. Other responses describing the use of an empty tin can to store things, such as coins or pens, are counted. Moreover, an answer is valid if the stated use is possible to implement, the respective item remains the main component of the new use and the implementation of the new use does not require further essential components other than simple resources which are typically available in an average household. 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 football, or a drum. In contrast, examples of invalid answers are a television, a computer, or a plane (unless specified as a toy). The validity score is determined by counting one point for each valid unusual use.

The second evaluation criteria is the flexibility score. It reflects the variety of responses of a subject and is determined by counting the number of different categories into which responses fall. Assume a subject gives the responses: candle holder, a baking-tin, a rattle, and a drum. The candle holder falls into the category decoration; the baking-tin falls into the category molds, and both the rattle and the drum are music instruments. That is, three different categories are mentioned. For each category, the participant receives one point resulting in a score of 3 flexibility points. Examples of frequently scored categories are "non-food containers" (e.g., pen container), "sport devices" (e.g., a football) or "toys" (e.g., pot hitting). The classification of categories is adopted from another experiment reported in Bradler et al. (2014). The latter tested about 600 subjects for the items tin can, paper and cord and developed a comprehensive catalog of responses and categories in which responses fall.<sup>11</sup> This comprehensive catalog of answers also gives precise information on how often particular answers are given.

This information is required for the third evaluation criteria: the originality of answers which is

---

<sup>9</sup>For instance, the correlation between subjects' scores for the item paper and tin can was 0.71, between paper and cord 0.72, and between tin can and cord 0.77.

<sup>10</sup>The original scoring procedures also use a fourth criterion, elaboration, which refers to the detailedness of the answers. This dimension is not used here to avoid that subjects spend too much time on very detailed answers, which would have made difficult a quick scoring of answers within the time frame of the experiment at risk.

<sup>11</sup>The Torrance Test of Creative Thinking offers an evaluation manual for the item tin can which was mainly developed with answers from people at relatively young age and restricted to English. Bradler et al. (2014) created a catalog of answers in German language and for the additional items cord and paper. Aligned to the original scoring manual for the tin can, approximately 55 categories were classified for each of the items.

measured by the statistical infrequency of an answer. All valid responses of previously tested subjects were tabulated according to their frequency and a rating scale to assess originality points was constructed. This scale allots one point to an answer if it is given by less than 5% of the test subjects (“original”) and two points to answers that were given by less than 1% of the test subjects (“very original”). Table 1 provides examples of frequent answers, categories as well as (very) original and invalid answers for all three items. In order to calculate the total score in the unusual uses task, the points for the separate dimensions validity, flexibility and originality are summed up. The scoring was conducted by three research assistants, who were carefully acquainted with the scoring procedures. In order to be able to evaluate the creative output quickly, sessions were restricted to 10 subjects. After each work task of 3 minutes, it took about 8-10 minutes to evaluate all answers. Subjects were informed that they were not allowed to use mobile phones or to read any private documents during these breaks. In comparison to expert ratings as an alternative evaluation method, the statistical approach of the scoring procedure in the unusual uses task substantially reduces the scope for subjectivity. Objective evaluation rules minimize the potential problem that the evaluation is perceived as arbitrary or unfair.

### 3.2 The Experimental Set-up

The aim of the experiment was (1) to observe how individual characteristics, and in particular creative ability, affect the sorting decision into a tournament-based vs. fixed pay scheme in a creative task and (2) to be able to compare results with previous findings by Dohmen and Falk (2011) who studied the latter for a simple math task. For this reason, the experimental design was largely adopted from Dohmen and Falk (2011). In total, subjects went through 11 steps in the experiment. Figure 1 gives an overview of all steps in the experimental design. The first five steps of the experiment were set up to measure individual creative ability and to provide subjects with feedback on their performance followed by some questions related to subjects’ effort and self-assessment. Creative ability was approximated by subjects’ productivity in the unusual uses task performed under two different conditions: with and without a monetary reward for performance. According to standard economic theory, subjects should increase effort until the marginal costs equal the marginal benefits of providing one additional unit of output. Thus, a linear payment per unit of output (a piece rate) should elicit a good proxy for individuals relative ability measured by their productivity. This may not hold for creative tasks if a crowding-out of intrinsic motivation occurs. Creative tasks have been found to involve a high level of intrinsic motivation relative to routine tasks (Amabile, 1996a), and therefore creative output may be relatively affected by financial rewards. This has empirically been observed in particular when a reward is perceived as controlling and decrease people feelings of self-determination (Deci et al., 1999). In contrast, when the informational value of a reward outweighs the controlling

effects, feelings of self-determination are not necessarily reduced but can also increase and lead to better creative performance (Eisenberger and Cameron, 1996; Eisenberger and Shanock, 2003). In economics, crowding-out effects have so far been studied mainly in the context of intrinsically motivated activities other than creative tasks (see Frey and Jegen, 2001, for an overview) and the existence of crowding-out effects is controversially debated (for instance, Fang and Gerhart, 2012; Fehr and Falk, 2002; Charness and Gneezy, 2009; Charness and Grieco, 2013). In order to check for potential crowding-out effects of the piece rate incentive, individual productivity is measured (1) without reward and (2) under a piece rate incentive.

In the following, I describe each step in more detail. The experiment started after subjects were instructed about the task and scoring procedure and the exercise period of one minute was completed. In step 1, each subject was asked to work on the unusual uses task with the item “paper”. In this first task, no payment was announced. In the following, subjects’ score in step 1 will be referred to as “Productivity Indicator 1”. As the performance in the first task was not rewarded, the Productivity Indicator 1 can be interpreted as a measure of intrinsic motivation and ability in the task.

After the task, subjects’ responses were scored as described in Section 3.1. Before feedback on scores was given, subjects were asked for a self-assessment of their absolute score in task 1 (step 2). In particular, subjects should indicate: “What do you think, how did you score in the previous task? (Please enter your guess for the total number of points for validity, flexibility and originality.)”.<sup>12</sup> This step represents an additional stage to the design of Dohmen and Falk (2011) and was implemented to measure how subjects self-assess their absolute performance before they receive any feedback on their actual score.

In step 3, subjects were informed about their actual total score and the following four questions were posed to elicit subjects perceived fun with the task and how hard they worked during the 3-minute working time: (1) “How much fun did you have during the task?” (2) “How much effort did you exert?” (3) “How stressed did you feel?” (4) “How exhausted did you get?”. Answers were given on a seven-point Likert scale, where the value 1 refers to “not at all” and the value of 7 to “very much”. Questions 2 to 4 were adopted by Dohmen and Falk (2011), question 1 was added as an indicator of intrinsic motivation.<sup>13</sup>

In step 4, subjects were asked to work on a second creative task. Before the task started, subjects were informed on the computer screen that they will face the same task as before but a different item, and this time they are paid a linear payment depending on their total score. Each point of their total score will be remunerated with 10 Taler (which was worth 0.20 Euros). The item

---

<sup>12</sup>Subjects were informed that they will receive a monetary bonus of 100 Taler if the answer is correct with a deviation of +/- 5 points.

<sup>13</sup>“Intrinsic motivation has been defined as performing behaviors out of interest, pleasure, and enjoyment.” (Waterman, 2005, p.1). People’s reports of how interesting and enjoyable they find the activity is an established measure of their intrinsic motivation (Vansteenkiste and Deci, 2003).

which each subject was given in the second task was a “cord”. In the following, subjects’ score in step 4 will be referred to as “Productivity Indicator 2”.

In step 5, subjects were asked again to subjectively assess the fun with the task and how hard they worked, identical to step 3. Subsequently, subjects were requested to indicate a self-assessment of their relative performance in the previous task by asking: “What do you think, how many of the 9 other participants achieved more points than you did?”.<sup>14</sup> Relative self-assessment will reveal whether subjects have a good knowledge about their relative performance which is an important factor for people’s sorting decision. A comparison of subjects’ true rank and their rank belief will reveal if subjects show an over- or underconfidence with respect to their creative productivity. The subjects received no information about their true rank in the distribution before the very end of the experiment.

In step 6, the subjects were informed that they have again the possibility to work for 3 minutes on the same task as before, but now they would be able to choose between two alternative compensation schemes. One compensation scheme was a fixed payment. In case the fixed payment was chosen, subjects would be guaranteed a payment of 400 Taler independent of their score.

$$w_i^F = 400. \quad (1)$$

The other compensation scheme was a tournament for a monetary prize. In case the tournament contract was chosen, a subject  $i$  competed against a randomly chosen opponent  $j$ , who had also opted for the tournament. Within this two-person tournament, the subject who achieved a higher score in the working period of 3 minutes won the tournament. The winner received a prize,  $\gamma$ , of 1300 Taler; the loser received 0 Taler. The wage of the tournament contract  $w^T$  of subject  $i$  can be written as follows:

$$w_i^T = \begin{cases} 1,300 & \text{if } x_i > x_j, i \neq j; \\ 1,300 \text{ with probability 0.5 and 0 with probability 0.5} & \text{if } x_i = x_j; \\ 0 & \text{otherwise} \end{cases}$$

Further details of the implementation were designed to mirror that of Dohmen and Falk (2011).<sup>15</sup>

---

<sup>14</sup>As in step 2, subjects received a monetary incentive to make their best guess. They received 100 Taler if the number was correct, 50 Taler for a deviation of +/- 1 from the correct number, and 0 Taler otherwise.

<sup>15</sup>Subjects were informed that in case of a tie, the winner would be determined by a random draw. If only one subject opted for the tournament (which did not happen), the subject was informed that the tournament could not take place and that the compensation would be the fixed payment of 400 Taler. In case of an odd number of subjects who opted for the tournament, the score of a randomly drawn subject was used a second time and assigned as an opponent for this unmatched subject.

Based on the framework in Lazear and Rosen (1981), the sorting decision can be briefly illustrated as follows. Assume an individual's output  $x_i$  depends on his productivity  $\pi_i$ , which is a function of effort  $e_i$  and ability  $\theta_i$ :  $x_i = \pi(\theta_i, e_i) + \epsilon_i$  where  $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ ,  $\pi_\theta, \pi_e > 0$  and  $\pi_{e\theta} \geq 0$ . Further, the utility of subject i is described by  $u(w, e) = w - c(e)$ , where w denotes the wage w and c the subject's costs which are a function of the given effort level e, with  $u_w > 0$ , and  $u_c < 0$  since  $c_e > 0$ . A utility-maximizing subject should opt for the tournament if the subject's expected utility of the tournament exceeds the subject's utility of the fixed pay. That is, a risk neutral subject optimally participates in the tournament if the tournament wage,  $w_i^T$ , multiplied by its probability minus its costs of effort exceeds the fixed wage,  $w_i^F$ , minus the subject's costs of a minimum effort provision,  $c(e^{min})$ , which can be interpreted as the costs of remaining in the lab:

$$w_i^T \cdot \text{Prob} \{ \pi_i(\theta_i, e_i^*) - \pi_j(\theta_j, e_j^*) > \epsilon_j - \epsilon_i \} - c(e_i^*) \geq w_i^F - c(e^{min}).$$

The probability of winning the tournament is determined by the difference in the productivity between subject i and subject j, which depends on their respective abilities,  $\theta_i$  and  $\theta_j$ , and the optimal effort levels,  $e_i^*$  and  $e_j^*$  compared to the difference in the random components  $\epsilon_j$  and  $\epsilon_i$ , which can be interpreted as luck.<sup>16</sup>

Because of the fact that the creative task is an established and validated creativity test, the difference in ability,  $\Delta\theta_{ij} = \theta_i - \theta_j$ , is likely to dominate luck in determining output, and thus the winner of the tournament. Thus, if subjects have a sufficient knowledge about their relative ability, one can expect a sorting pattern where more able (creative) subjects are more likely to sort into the tournament while less able (creative) subjects opt for the fixed wage. However, due to the specific attributes of the creative task, which I discussed in the introduction, subjects may not assess their own ability relative to the ability of the other subject correctly. Subject's belief on the difference in ability,  $\hat{\Delta\theta}_{ij} = \hat{\theta}_i - \hat{\theta}_j$ , might be optimistic with  $\hat{\Delta\theta}_{ij} > \Delta\theta_{ij}$ , or pessimistic with  $\hat{\Delta\theta}_{ij} < \Delta\theta_{ij}$ . An optimistic (or pessimistic) subject might still prefer the option which would have been favored with a correct assessment of own relative ability. However, subjects may be overly optimistic (or pessimistic) so that they favor the tournament (the fixed pay) when a correct assessment of own relative ability would not have led to the same decision. That is, misjudgments may render subjects less likely to sort into the tournament according to their true relative ability in the creative task.

---

<sup>16</sup>As Dohmen and Falk (2011) already noted, there is not necessarily a unique sorting threshold as this depends on various distributional assumptions. For instance, in the absence of luck, a more able subject always wins the tournament against a less able subject. In this case, no tournament would take place as only the most able subject would opt for the tournament. In contrast, if, for instance, both agents have the same ability, luck becomes the decisive factor and everybody should opt for the tournament (in case of risk neutral agents). However, it is plausible to assume that a sorting equilibrium exists for intermediate cases where subjects opt for the tournament if their ability exceeds a certain threshold and less able subjects sort into the fixed payment.

In step 7, immediately after opting for one of the two payment schemes but before the working time started, the subject faced a series of hypothetical sorting decisions. Here, subjects were asked which scheme they would have picked if the fixed payment had been different. More precisely, subjects who preferred the tournament could decide whether they still would have preferred the tournament over a fixed payment of {450, 500, ..., 800}. If the fixed payment was preferred in the actual sorting decision, subjects should indicate whether they still would have decided against the tournament in case of a fixed payment of {50, 100, ...350}. These hypothetical sorting decisions reveal how strong subjects' preferences for the respective payment schemes are.

Subsequently, the actual working period of 3 minutes started in step 8 under the preferred pay scheme. After the working time ended, the subject's earnings were displayed on the screen, and tournament participants were notified about their score and the score of their randomly matched opponent. Identical to step 3 and 5, subjects are again asked to assess their fun with the task and how hard they worked (step 9). The final two steps are designed to observe personal characteristics.<sup>17</sup>

In step 10, risk preferences were elicited with simple lottery choices. Subjects faced a sequence of 15 choices between a safe option and a lottery. The lottery was the same in each row and offered 400 Taler with 0.5 probability and 0 Taler with 0.5 probability. The safe option guaranteed a fixed amount of Taler and got increasingly attractive from row to row. In row 1, for instance, the safe option was 25 Taler, in the second 50 Taler and so on up to 375 Taler as alternative to the lottery. In order to determine subject's payoff, one row was randomly chosen to be relevant.<sup>18</sup> In the final step 11, subjects received a questionnaire to gather information on subjects' socio-economic characteristics (such as gender, age, nationality and income), their educational background and achievements (final school grade, major subjects in high school and field of study at university), creative hobbies as well as the Big Five personality traits (Costa and McCrae, 1992). The questionnaire further included an alternative measure for participants' risk attitudes. Like Dohmen and Falk (2011), I used the wording from the 2004 wave of the German Socio-Economic Panel Study (GSOEP).<sup>19</sup> Here, subjects are asked to indicate their

---

<sup>17</sup>In contrast to Dohmen and Falk (2011), I did not measure social preferences because the selection into a tournament in their experiment was neither driven by subjects' willingness to trust nor their inclination to reciprocate. Also, Bradler et al. (2014) find that for subjects who were randomly selected into a tournament in the unusual uses task neither reciprocal preferences in a gift-exchange game nor their revealed social preferences in a sequence of simple distribution games showed a significant impact on subjects' effort levels in the tournament.

<sup>18</sup>The fact that subjects already know their earnings from previous stages might cause an endowment effect and bias subjects' preferences in the lottery choices. Especially, the information on the assignment of the relatively high tournament prize can be seen as critical to cause an endowment effect. In order to check for the validity of the lottery risk measure., I additionally randomized the information structure across sessions: in half of the sessions subjects who performed in the tournament learned their payoff immediately after the task, and in the other half subjects were not informed about the outcome of the tournament until the very end of the experiment.

<sup>19</sup>The German Socio-Economic Panel (GSOEP) is a representative large-scale survey among the German

willingness to take risks in general on an 11-point scale where zero means “not willing to take risks at all” and 10 means “strongly willing to take risks”. The behavioral validity of this survey measure was demonstrated by Dohmen et al. (2011) who compare responses for this risk measure with behavior in paid real-stakes lotteries for a large representative sample. The risk measure from the GSOEP proved a high predictive power for actual risk-taking behavior in the financial domain, such as holding stocks, but was also the best all-around measure for risky behavior across other contexts such as health, sports or labor market outcomes. In addition, I elicited subjects’ intrinsic and extrinsic motivational orientation with the help of the Work Preference Inventory (WPI) invented by Amabile et al. (1994).<sup>20</sup>

### 3.3 Procedural Details

The experiment was computerized using the software z-Tree (Fischbacher, 2007) and conducted with students from the University of Heidelberg. Subjects were recruited using the online recruitment system ORSEE (Greiner, 2004). At the very beginning of the experiment, subjects were instructed that no aid was allowed to come up with creative ideas in the unusual uses task (no mobile, no internet or any private documents). They were also informed about the basic structure of the experiment and that all actions during the experiment were anonymous. Subjects received a printed handout with an example of the unusual uses task and the scoring procedure. The latter was explained with examples for the item “old tire”. Further instructions were shown on the computer screen. In total, I ran 13 sessions, each with 10 invited participants. Thus, overall 130 students participated in the experiment.<sup>21</sup> A session lasted on average 90 minutes. The exchange rate of the experimental currency “Taler” was 10 Taler = 0.20 Euros (1 Euro = 1.32 US Dollar at that time). Average earnings were 21.30 Euros.

---

population covering a broad range of topics.

<sup>20</sup>Amabile developed a measure to identify the motivational orientation type. The measure consists of 30 items describing personal attitudes and characteristics as for example: “To me, success means doing better than other people.” (see Amabile et al. (1994), p. 956). People are asked to agree or disagree to the statements on 4-point Likert scales. Finally, a score for the extrinsic and intrinsic motivational orientation can be constructed. Extrinsic motivation types are assumed to be more responsive to financial incentives than intrinsic motivation types. Therefore, the WPI measure gives the possibility of observing whether the tournament prize attracts people with relatively high extrinsic motivation.

<sup>21</sup>Fluent German skills were announced as a prerequisite for participating in the experiment. Nevertheless, six students had an insufficient knowledge of the German language, and thus, severe problems to understand the task. Another subject did not press “Enter” after submitting responses, so that creativity ability could not be measured. These seven participants were dropped for the analysis.

## 4 Results

In the results section, I will first look at descriptive statistics on the creative productivity of subjects before and after sorting. Subsequently, I will analyze how subjects sorted into their preferred compensation scheme according to their creative productivity, self-assessments, risk preferences, gender, and other personal characteristics.

### 4.1 Creative Productivity before Sorting

Individual productivity was measured in the first task of the experiment without any reward (Productivity Indicator 1) and in the second task with the financial incentive induced by a piece rate scheme (Productivity Indicator 2). On average subjects achieved a total score of 22.57 (s.d.=10.13) in the Productivity Indicator 1 and 27.07 (s.d.=10.26) in the Productivity Indicator 2.<sup>22</sup>

The difference is statistically significant at the 1% level in a paired t-test. This suggests a positive incentive effect of the piece rate on creative performance if the assumption holds that subjects were not affected by learning effects. The latter is supported by the evidence of Bradler et al. (2014) where subjects who performed the unusual uses task with identical items and work duration did not show significant learning effects. Thus, there is no indication of a detrimental effect of the piece rate on creative performance in the unusual uses task and the Productivity Indicator 2 appears to be a good proxy for individual ability.<sup>23</sup> Self-reports on perceived fun on the task provide additional support against the notion of a potential crowding-out effect of intrinsic motivation by the piece rate. If crowding-out occurred, one should see a decline in perceived fun on the task under the piece rate compared to perceived fun on the task in the absence of rewards. The opposite is true. Subjects' indicated an average fun level of 3.53 (s.d.=0.13) after task 1 (no reward) and 4.00 (s.d.=0.14) after task 2 (rewarded by piece rate). The difference is significant by a paired t-test ( $p=0.000$ ) and demonstrates that the piece rate

---

<sup>22</sup>After completing the first creative task, subjects were asked in step 2 of the experiment to give a guess how many points they achieved in the first task. On average, subjects underestimated their absolute score by 10.28 points (s.d.=14.85). This demonstrates that subjects had difficulties assessing their absolute performance. In order to enable subjects to learn about their creative ability in absolute terms, subjects received feedback on their score after task 1 and 2 before they faced the sorting decision. However, what finally determines whether a subject wins or loses the tournament is relative performance.

<sup>23</sup>As described in the introduction, a financial reward for a creative task may reduce performance due to a crowding-out of intrinsic motivation (Deci et al., 1999; Frey and Jegen, 2001). However, previous psychological literature argues that crowding-out effects are largely limited to rewards which are perceived as controlling Deci et al. (2001) whereas an enhancing effect occurs if the latter instead strengthens subjects' feeling of self-determination (Eisenberger and Aselage, 2009; Eisenberger and Armeli, 1997; Eisenberger and Cameron, 1996).

increased intrinsic motivation in the task.<sup>24</sup> Moreover, a comparison of perceived stress after task 1 and 2 reveals no significant change between both tasks.

In the following, I will mainly refer to Productivity Indicator 2 in the analyses, but robustness checks reveal that all main results stay qualitatively similar if the Productivity Indicator 1 is used instead.

The data gathered in the experiment also shed light on the relationship between creative productivity and personal characteristics such as risk attitude, gender, age, the high school grade or the Big Five personality traits. In line with comprehensive evidence in the literature (for instance, Batey and Furnham, 2006), I find a substantial and significant correlation between creative performance and intelligence where the latter is approximated by the final grade from high school.<sup>25</sup> Another significant and sizable correlation can be found for creative performance and the Big Five personality trait openness (Spearman rank correlation, rho=0.4335, p=0.000) which is also an established finding in the literature (for instance, McCrae, 1987).

## 4.2 Creative Productivity after Sorting

In the final creative task, subjects could choose their preferred reward scheme. According to principal-agent-theory, the output of subjects who chose the tournament is supposed to reveal a higher creative performance compared to the output of subjects who chose the fixed payment for two reasons: (1) the tournament prize should trigger a positive incentive effect and (2) the more productive agents should self-select into the tournament.

Before I will analyze whether agents sorted according to their productivity, I investigate differences in creative output of subjects before and after sorting. By comparing average productivity, subjects in the tournament outperformed subjects under fixed pay with a considerable performance gap: 35.853 (s.d.=11.849) vs. 23.763 (s.d.=11.589) points, see Table 2. The subjects who opted for the tournament significantly increased their score under the tournament compared to their creative productivity in tasks 1 and 2 (paired t-test, p<0.000 for both comparisons). When performance is compared for the same subjects over time, one can argue that these performance changes are due to learning effects which cannot be excluded here. However, another study by Bradler et al. (2014) used a very similar set-up with identical items and time frame where subjects in a control group did not show any significant learning effects in

---

<sup>24</sup>This is in line with empirical evidence for a crowding-in of intrinsic motivation, as for instance, provided by Eisenberger and Cameron (1996) or Eisenberger et al. (1999). Robert Eisenberger and his co-authors argue that crowding-out effects are more likely for performance-independent rewards as they signal that the principal rewards average or inadequate performance. In contrast, performance-dependent rewards such as a piece rate rather foster people's feeling of self-determination, and thus their creative performance.

<sup>25</sup>The final grade from high school refers in Germany to the university-entrance examination (Abitur). Spearman rank correlation of Productivity Indicator 2 and the final grade is -0.23 (p=0.011), where a grade of 1 refers to the best and a grade of 4 refers to the worst grade which can be assigned.

the unusual uses task. This suggests that those individuals who opted for the tournament show a positive incentive effect under the tournament compared to the piece rate, and an even stronger incentive effect compared to the performance under no reward. This positive incentive effect is in contrast to studies showing crowding-out effects of extrinsic motivators in tasks which are high in intrinsic motivation, such as creative tasks, which was controversially discussed in the psychological literature (see, for instance, Deci et al., 1999; Amabile, 1996a; Eisenberger and Cameron, 1996). However, the positive incentive effect is in line with Bradler et al. (2014), who also find a strong incentive effect of a randomly assigned tournament in the unusual uses task. Subjects who preferred the fixed payment show a creative output after sorting which is slightly, albeit not significantly, larger than in task 1 without reward (paired t-test,  $p=0.14$ ) but somewhat lower than in task 2 under the piece rate which is marginally significant (paired t-test,  $p=0.09$ ).

### 4.3 Sorting into the Tournament

#### The impact of individual creative productivity

In this section, I will investigate to what extent the difference in performance between the tournament and the fixed payment scheme goes beyond an incentive effect and can also be attributed to a sorting effect. Therefore, I analyze whether the subjects who were relatively productive under the piece rate (Productivity Indicator 2) were more likely to sort into the tournament. In total, 55 subjects chose the fixed payment in the sorting decision and 68 subjects chose the tournament. Figure 2 illustrates the cumulative distribution of creative performance measured by Productivity Indicator 2 for subjects who opted for the tournament and those who opted for the fixed payment, separately. If subjects sort into the pay schemes according to their productivity, the distribution functions should be clearly separated from each other, in the way that the output distribution of subjects opting for the tournament should largely dominate the output distribution of subjects who sorted into the fixed-pay scheme. In contrast, they run very close to each other and at some points in the distribution they even cross. This evidence suggests that no systematic productivity sorting of agents in the two payment schemes exists.

Table 2 confirms this notion in another way. The latter reports means and standard deviations for both productivity indicators separately for subjects who later chose the tournament and those opting for the fixed payment. On average, subjects who sorted into the tournament achieved 27.824 ( $s.d.=9.638$ ) points in Productivity Indicator 2. In comparison, subjects who later opted for the fixed payment, scored 26.145 ( $s.d.=10.997$ ) points. This difference is not statistically significantly different by a t-test ( $p=0.370$ ). The result also holds for the scores of the Productivity Indicator 1. Here, a t-test yields a p-value of 0.507.<sup>26</sup>

---

<sup>26</sup>The result is also robust to a comparison test with subjects' maximum productivity out of both productivity indicators.

It is also interesting how subjects sorting decisions are related to their scores in the separate dimensions of their total score: validity, flexibility and originality. It might be the case that the tournament attracted subjects who do not submit a high quantity of answers but a high quality in terms of the originality of answers. Figure 3 demonstrates that the separate scores (Productivity Indicator 2) of subjects who opted for the tournament were not significantly better than were the scores of subjects preferring the fixed payment in all scoring dimensions (Wilcoxon rank-sum tests, p-values are 0.18 for validity, 0.27 for flexibility and 0.43 for originality).

In step 6 of the experiment, subjects were asked for their hypothetical sorting decision for alternative fixed payments. One should expect that the more attractive an alternative fixed payment is, the fewer workers should prefer to select into the tournament. This is borne out by the data. Figure 4 (bottom part) illustrates a declining fraction of subjects who prefer the tournament when the offered fixed payment is increasing. Furthermore, the higher the offered fixed payments, the more productive are the subjects who still prefer the tournament. However, a Spearman rank correlation reveals that the relationship between subjects' productivity and the lowest fixed payment they prefer over the tournament is not significant ( $p=0.434$ ). Figure 4 (upper part) shows that only in the range of low fixed payments are some small differences in the individual productivity of subjects sorting into the tournament and subjects opting for the fixed payment. In the higher range of fixed payments, subjects' preferences in favor of or against the tournament do not relate to their productivity.

In sum, the increase in creative productivity in the tournament seems to be driven only by an incentive effect. There is no indication of a significant selection effect of the more creative subjects into the tournament.

### **The impact of self-assessment and personal characteristics**

As mentioned in the introduction, other potential drivers apart from creative ability may affect agents' self-selection. I will first present descriptive statistics on the sorting of subjects with respect to their risk attitudes, relative self-assessments, gender, as well as other self-reported variables and Big Five personality traits. Finally, I present results of a probit regression on subjects' sorting behavior which takes all relevant sorting dimensions simultaneously into account.

The risk attitude is likely to affect the sorting decision because the two payment schemes offered differ in the risk they involve. In the case of the tournament, subjects face an uncertain payoff because of a large spread in potential earnings (1,300 vs. 0 Taler) and because the opponents' performance is unknown. In contrast, in the case of the fixed payment agents always receive 400 Taler independent of their output; no risk is involved. The more risk averse subjects are, the lower are their expected utilities from the tournament prize. Thus, it can be expected that the more risk averse a subject is, the less likely he or she is to select into the tournament. This

prediction is borne out by the data. In the following, I focus on the risk measure elicited in step 10 of the experiment (the validated question from the German socio-economic panel) as Dohmen and Falk do for their analysis.<sup>27</sup> Subjects who opted for the tournament exhibit on average a risk attitude of 5.91 (s.d.=2.11) while subjects choosing the fixed payment reveal an average of 4.65 (s.d.=1.98) on an 11-point scale (where zero means “not willing to take risks at all” and 10 means “strongly willing to take risks”). This difference is significant by a Wilcoxon rank-sum test ( $p=0.001$ ). In order to compare the influence of risk preferences on the sorting decision to the results from Dohmen and Falk (2011), I also analyzed whether subjects who chose the tournament in the creative task exhibited a different risk tolerance than those who favored the tournament for the math task. As mentioned in the introduction, creative tasks are characterized by the fact that no logic algorithm or standard approach exists to generate creative ideas. As a consequence, the absence of any significant productivity sorting might also be driven by the fact that subjects perceive creative output as per se less predictable and more variable than performance in a math task. If this were the case, subjects who opt for the tournament in the creative task should reveal a higher risk tolerance compared to subjects favoring the tournament in Dohmen and Falk (2011). This is not borne out by the data. The average willingness to take risks of subjects who choose the tournament is 6.02 (s.d.=2.00) in Dohmen and Falk (2011) and 5.91 (s.d.=2.11) in the present study, which does not significantly differ (Wilcoxon rank-sum test,  $p=0.917$ ).<sup>28</sup>

Hence, the absence of a sorting pattern with respect to creative productivity cannot be explained by a larger (perceived) risk that creative output will not correspond to perceived ability and effort. Nevertheless, the descriptive results suggest that risk preferences significantly affect agents’ sorting into a tournament scheme.

As explained in Section 3.2, subjects were asked for a relative self-assessment of their performance in step 5 of the experiment. Given the creative productivity of a subject, it can be expected that the more optimistic a subject’s relative self-assessment is, the more likely

---

<sup>27</sup>Like Dohmen and Falk (2011), I observe a few subjects who indicate multiple switching points in the risk measure with lottery choices (step 9), where it is not clear how to treat those observations. Further, the lottery risk measure might be biased by endowment effects due to previous earnings in the experiment. As described in 3.2, I randomized half of the subjects into sessions with immediate information about tournament winners and losers (as in the original design by Dohmen and Falk, 2011) and the other half into sessions where this information is provided at the very end of the experiment to check whether the information on earnings affect behavior in the lottery choices. The results indicate that this is the case. For instance, tournament participants who do not know whether they won the tournament or not are significantly more risk tolerant in the lottery choices than subjects who already received the information of being the winner (Wilcoxon rank-sum test,  $p=0.068$ ). No such difference can be found for the GSOEP risk measure (Wilcoxon rank-sum test,  $p=0.986$ ). A similar pattern can be found for losers. Thus, the GSOEP risk measure appears to be the more reliable indicator of risk preferences in this design.

<sup>28</sup>The share of people opting for the tournament in the present study is also similar to that in Dohmen and Falk (2011): 55% in this study vs. 50% in Dohmen and Falk (2011). Further, the average willingness to take risks of all subjects in the experiment does not differ significantly between Dohmen and Falk (2011) and the present study (5.24 with a standard deviation of 2.18 in Dohmen and Falk (2011) vs. 5.35 with a standard deviation of 2.14 in the present study, Wilcoxon rank-sum test,  $p=0.600$ ).

the subject self-selects into the tournament. This is supported by the data. Subjects who later preferred the tournament guessed on average a rank of 5.06 (s.d.=1.53) while those subjects who later preferred the fixed payment were more pessimistic by assuming on average a rank of 6.09 (s.d.=1.70) among the 10 participants within their session. This difference is statistically significant by a Wilcoxon rank-sum test ( $p=0.001$ ). If subjects have a very good knowledge about their own relative ability, these self-assessments should simply reflect revealed productivity. That is, self-assessed ranks should be very highly correlated with the Productivity Indicator 2. As Table 3 shows, the correlation is significant but with a magnitude of -0.3541 ( $p=0.001$ ) substantially weaker than in Dohmen and Falk (2011) where productivity and self-assessments correlated with a magnitude of -0.695.<sup>29</sup>

In order to compare how subjects evaluated their own relative performance and how they truly performed, one can look at rank beliefs and true ranks. If subjects guess their relative rank correctly, rank belief and true rank are identical. On average, the absolute difference between subjects' rank belief and their true rank is 2.22 (s.d.=1.61). Figure 5 illustrates the individual differences between rank beliefs and true ranks graphically.

The latter demonstrates substantial misjudgments of subjects. Only 16 out of 123 subjects (13.01%) predict their rank correctly. A total of 44.72% of the subjects misjudged their rank by indicating a worse performance rank than their true rank, thus, these subjects can be referred to as underconfident. The remaining 42.28% indicated a better rank than they actually achieved and can be described as overconfident. Moreover, one can look at the percentage of subjects who misjudged their rank only by one rank position or more than this. A total of 26.83% of the subjects misjudged their true rank by one rank position and were either underconfident (8.94%) or overconfident with respect to their rank (17.89%). Further 35.77% of the subjects underestimated their relative performance by more than one rank position and the remaining 24.39% estimated their relative performance too optimistically by guessing more than one rank better than they actually achieved. This suggests that subjects had difficulties correctly assessing how well they performed relatively to others in the creative task. The misjudgments in self-assessments go in both directions: a similar frequency of subjects showed to be over- or underconfident. This picture changes if self-assessments are split up by gender. Among men 50% of all subjects are overconfident, whereas only approx. 35% of all women are overconfident. However, self-assessments are not significantly different between men and women (Wilcoxon rank-sum test,  $p=0.25$ ).

As mentioned in the introduction, many studies have found evidence that women are less likely to enter competitive environments (for instance, Niederle and Vesterlund, 2007). Differences

---

<sup>29</sup>Please note, that the negative sign of the correlation is due to the measurement of people's self-assessments. People where asked to indicate their belief on the number of better performing participants. That is, the larger the indicated number, the more pessimistic is the self-assessment.

in competition aversion between men and women have been related to women being more risk averse (for instance, Eckel and Grossman, 2008; Dohmen and Falk, 2011) or to the stereotype that women are less able in specific tasks such as a math task, although this is in fact not the case (Günther et al., 2010). No such stereotype should exist for the "unusual uses task". Female subjects indicated an average willingness to take risks of 5.44 (s.d.=2.22) and male subjects an average of 5.24 (s.d.=2.06) which is not statistically different (Wilcoxon rank-sum test,  $p=0.663$ ). And looking at the raw data, the share of men and women choosing the tournament is very similar: 53.85% of all women and 56.9% of all men opted for the tournament. Thus, the descriptive data do not indicate a gender effect in the sorting pattern.

After each task in the experiment, self-reported variables on perceived fun, effort, stress and exhaustion were elicited. Table 4 presents means and standard deviations for all self-reported variables before and after sorting by the preferred compensation scheme.<sup>30</sup> Before sorting, effort and exhaustion do not differ significantly between subjects preferring the tournament and those opting for the fixed payment. In contrast, Table 4 shows that subjects who sort into the tournament indicated a significantly higher level of perceived fun and a significantly lower level of stress in the previous task than did subjects who sorted into the fixed payment.<sup>31</sup> This does not suggest that highly intrinsically motivated subjects anticipated a detrimental effect of the tournament on performance. Based on a comprehensive literature (for instance, Amabile, 1996a), fun as a proxy for intrinsic motivation can be assumed to be positively related to creative performance. In contrast, stress has been found to be negatively related to creative performance (for instance, Avey et al., 2012). Spearman rank correlation coefficients in Table 3 confirm that the perceived level of fun is positively and stress is negatively related to the Productivity Indicator 2. However, the magnitude of the correlation coefficients are not large enough to mediate a significant relationship of creative performance and sorting into the tournament (Spearman rank correlation of Productivity Indicator 2 with fun:  $\rho=0.3445$ ,  $p<0.001$ ; and with stress:  $\rho=0.157$ ,  $p=0.083$ ).

Interestingly, women experience significantly more fun and exhaustion than men at the creative task (Wilcoxon rank-sum tests,  $p=0.002$  for fun, and  $p=0.03$  for exhaustion). Finally, neither the Big Five Personality traits nor the Work Preference Inventory (WPI) for general extrinsic and intrinsic motivation (Amabile et al., 1994) showed a significant difference between subjects opting for the tournament and those opting for the fixed payment (Wilcoxon rank-sum tests for Big Five,  $p>0.15$  for all traits, and  $p>0.36$  for the WPI scales).

---

<sup>30</sup>In the following, I will focus on self-reported measures for performance under the monetary incentive of the piece rate (step 4) as this seems most informative for the selection into a tournament for a monetary prize. However, results are very similar for self-reported measures which were elicited after the first creative task where performance was not rewarded (step 2).

<sup>31</sup>The same is true for mean comparison tests of self-reports elicited in step 2: subjects who later choose the fixed payment indicated significantly less fun ( $p=0.007$ ) and significantly more stress ( $p=0.074$ ) than did subjects who chose the tournament.

## **Joint analysis of multidimensional sorting**

In the following, I will present a multivariate regression which takes all important sorting dimensions simultaneously into account, and analyzes their relative importance for the sorting decision. A probit model is estimated where the dependent variable takes the value one if a subject selected into the tournament, and zero otherwise. The reported coefficients are marginal effects estimated at the mean. They can be interpreted as the change in the probability that a subject opts for the tournament when the respective variable increases by one unit at the mean. Table 5 reports the results. Column II adds the self-reported measures fun, stress, effort and exhaustion (elicited in step 5). Columns III and IV add control variables such as age, income, the Big Five personality measures and a dummy for subjects whose field of study is economics.

For the interpretation of the results, I will mainly refer to the most comprehensive specification in column IV. The first important result that the analysis reveals is that individuals' creative ability, as measured by the Productivity Indicator 2 does not explain sorting behavior. This further strengthens the result from the previously presented descriptive statistics.<sup>32</sup> Again in line with the descriptive statistics, regression results demonstrate that instead of ability, risk preferences and relative self-assessments explain a large share of the sorting pattern. If a subject indicates a risk tolerance which is one point higher than the average indicated willingness to take risks on an eleven-point scale the likelihood increases by 7.6 percentage points that this subject opts for the tournament. The magnitude of the coefficient for risk attitude is similar but slightly lower than in the analysis of Dohmen and Falk (2011) who find an increase in the likelihood to choose the tournament of 8.7 percentage points for a one unit increase of the willingness to take risks at the mean. A subject's perception of the number of better performing participants within the session is significantly negatively related to the choice for the tournament. That is, the lower the value of the relative self-assessment, the higher is the rank a subject believes to have among the total of 10 participants within a session. Subjects with a more optimistic self-assessment of one rank below the mean are by 8.4 percentage points more likely to enter the tournament (and vice versa for pessimistic subjects). The magnitude of the coefficient is substantially larger than in Dohmen and Falk (2011), who find an increase of 2.7 percentage points. This indicates that subjects do strongly rely on their self-assessments when facing the sorting decision.

Self-reports on perceived fun, effort and stress after performing under the piece rate (step 5) did also affect the likelihood of entering the tournament. It turns out that perceived fun as well as subjects' indicated effort level (on a seven-point Likert scale) show a strong positive effect on the likelihood to enter the tournament. In contrast, perceived stress is strongly negatively related to subjects choice to opt for the tournament.

---

<sup>32</sup>Robustness tests with Productivity Indicator 1 or specifications with different subsets of the independent variables show very similar results.

Finally, I would like to look at whether sorting is affected by gender. As shown earlier with the raw data, the fraction of men and women entering the tournament was very similar. In columns I and III, where individual productivity, relative self-assessments and gender are jointly taken into account, no significant effect of women for the entry into the tournament emerges. Before concluding that sorting is not affected by gender, I turn again to Table 3, which shows that females report significantly higher level of fun and exhaustion in the creative task.<sup>33</sup> Because fun turned out to be a significant positive predictor of favoring the tournament, it might be the case that a potential gender effect is confounded by this variable. Columns II and IV report regression results where individuals perceived level of fun, effort, stress and exhaustion are additionally included as explanatory variables. Controlling for perceived fun, it turns out that women are less likely to enter the tournament. This is in line with prior evidence that women are averse to competition which goes beyond mere risk aversion (Niederle and Vesterlund, 2007).

## 5 Supplementary Analysis with Representative Field Data

### 5.1 Data and Design

A legitimate question is whether the experimental results can be generalized to real labor market settings. An attractive approach which was exploited by Dohmen and Falk (2011) is a complementary analysis of a large field data set as an external robustness check. This may not offer the possibility of excluding potential biases due to unobserved heterogeneity or reverse causality. But if similar sorting patterns emerge in real-world data, the complementarity of the results suggests that the systematic pattern which was found in the lab also transfers to real world labor market. In order to test for the external validity of my results, I conduct a complementary analysis with a representative field data set from Germany which allows a comparison of sorting into performance pay between creative and non-creative jobs.

Like Dohmen and Falk (2011), I use data from the German Socio-Economic Panel (GSOEP) which is a representative longitudinal study of private households in Germany. The survey targets all household members who are more than 17 years old, and most of the interviews are conducted either face-to-face or via a personal phone interview.<sup>34</sup> Respondents are surveyed according to a wide range of topics, including their current employment status, education, and various personal characteristics. In selected years, the questionnaire included a question on whether job performance is regularly assessed by a superior in a formal procedure. Dohmen

---

<sup>33</sup>In Table 3 correlations are presented for the self-reported measures of fun, effort, stress and exhaustion with respect to the second creative task (step 4). However, correlations reveal a very similar picture if self-reports after the first creative tasks (step 2) are used, where subjects were not rewarded.

<sup>34</sup>For further information on the GSOEP data, see <http://www.diw.de/en/soep>.

and Falk (2011) use this information as a rough indicator for variable pay and estimate how productivity, risk attitudes, social preferences and gender affect the probability of working under variable pay.<sup>35</sup> As proxy variables for productivity, they use individuals' years of schooling, experience on the labor market and tenure. The picture which emerges is very similar to the laboratory results. They find that more productive and risk tolerant workers are more likely to work for performance pay while women are less likely to work under variable payment schemes. They conclude that these findings are consistent with the observed sorting behavior in their experiment.<sup>36</sup>

In the following, I present results from an analysis of the GSOEP data to check whether sorting into performance pay differs between creative and non-creative jobs and whether my experimental results can be found in a similar manner in the GSOEP data. The present analysis extends the evidence on sorting into performance pay provided by Dohmen and Falk (2011) in several respects: the analysis is extended with data from further waves, different variations of the dependent variable are presented, and most importantly, a comparison between creative and non-creative jobs is added. As self-assessments of creative performance are not available in the GSOEP, the analysis focuses on the impact of creative ability on sorting into performance pay, and further takes other relevant sorting dimensions such as additional productivity indicators, risk attitude, gender, and other firm characteristics and occupational categories into account.

A crucial requirement for the analysis is the definition of creative jobs. Therefore, an established classification suggested by Richard Florida (2004, 2006) is adopted. Similar to the concept of creativity applied in the experiment, Florida's notion of the creative class corresponds to a relatively broad definition. He states that the distinguishing characteristic of the creative class is that its members "engage in work whose function is to create meaningful new forms" (Florida, 2002, p. 68). Florida distinguishes those who fully engage in the creative process and whose primary job function is to be creative and innovative as the creative core. An important subgroup of the creative core comprise the Bohemians, who are artistically creative people such as writers, painters, designers, photographers, journalists, dancers, composers, actors and musicians. Apart from this small subgroup, the creative core comprises a wide range of occupations in the fields of science, engineering, education, computer programming and research.<sup>37</sup> According to Florida, the members of the creative core are those who mainly engage in creating something new: ideas,

---

<sup>35</sup>About 80% of respondents report that performance evaluations affect their monthly gross wage, a yearly bonus, future salary increases and/or potential promotions. The question on risk preferences is explained in Section 3.2 because the identical question was used in step 11 of the experiment to elicit subjects risk preferences.

<sup>36</sup>Grund and Sliwka (2010) conducted a related study with GSOEP data and showed similar results.

<sup>37</sup>Beyond the creative core, Florida (2004) defines a group of creative professionals: a diverse group of professionals who "engage in creative problem-solving, drawing on complex bodies of knowledge to solve specific problems." and are required to regularly think on their own. The creative professionals include all kind of senior officials and managers, health professionals such as nurses, finance or sales professionals. The share of core creative activities such as idea generation can, however, be seen as relatively small in these occupations. That is why I restrict my analysis to the Bohemians and the creative core.

knowledge, technology, designs and/or content. Divergent thinking is required here in a similar way as it is the case in the unusual uses task. An overview of all of the occupations that I include in the creative class can be found in Table 6.

In contrast to Dohmen and Falk (2011), who use the 2004 wave, I extend the empirical analysis of the GSOEP to the waves 2008 and 2011 where the relevant variables have also been elicited.<sup>38</sup> Over the time span from 2004 to 2011, both sectors show an substantial increase in the number of employees with pay for performance. In creative jobs, the number of workers who indicated to be regularly evaluated according to their performance increased from 35% to 52%, and in non-creative jobs it shows a rise from 28% to 37%. In the GSOEP subjects were further asked whether this performance evaluation has some monetary impact, by affecting their monthly gross wage, a yearly bonus, future salary increases and/or potential promotions. In both sectors, on average approximately 80% of respondents indicated that they experience at least one of these monetary impacts. The portion of respondents that experience direct monetary consequences, such as an impact on monthly gross wages or a yearly bonus, increased from 15% to 26% for the creative class and from 14% to 20% for the non-creative class. Hence, although performance evaluations are more common for the creative class, the relative frequency of performance pay in terms of direct monetary consequences on monthly income or yearly bonuses is very similar in both sectors.

Given its broad application, not much is known about the extent to which performance pay does not only serve as a tool to motivate workers but also attracts the most creative talents in jobs which require creative output. As a proxy for creative productivity, I use two indicators: the years of schooling as an indicator of intelligence and the Big Five personality trait openness. An extensive research has shown that creativity is strongly related to intelligence (Eysenck, 1995; Batey and Furnham, 2006). Thus, years of schooling, as a rough proxy for intelligence, should be not only related to an individual's general productivity but also to an individuals' capacity to be creative. Yet, empirical evidence indicated that the relation is not linear and gets weaker for IQ's greater than 120 (see Batey and Furnham, 2006, and references therein). As a large share of employees in the creative class is highly educated, the years of schooling might not serve as a good predictor for these individuals.<sup>39</sup> That is why I use the Big Five personality trait "Openness to experience" as an additional indicator for creative productivity. The Big Five model is an established concept of describing human personality with five factors: conscientiousness, extraversion, agreeableness, openness, and neuroticism (Costa and McCrae, 1992). The importance of personality traits for creative outcomes has received much attention

---

<sup>38</sup>The three waves are merged for a cross-sectional analysis. If an individual is observed in more than one year, the most recent observation is used.

<sup>39</sup>Employees in the creative sector have a larger mean of the years of schooling than employees in the non-creative sector. However, the variance of the years of schooling does not differ between the creative and non-creative class (variance comparison test,  $p>0.75$ ).

by modern creativity research, and the personality trait openness has revealed the strongest and most robust relation to various dimensions of creativity, including divergent thinking.<sup>40</sup> Openness to experience describes a cognitive style which distinguishes individuals who are imaginative, curious, amenable to new ideas, experiences and unconventional perspectives from those who prefer the conventional, routine, and familiar (Costa and McCrae, 1992). For the measure “Openness to experience”, participants in the GSOEP have been asked to indicate on short scales whether they have a broad imagination, like to contribute their own ideas, and appreciates artistic or esthetic experiences.<sup>41</sup> Employees with a higher score in openness have been observed to score higher in creativity tests (McCrae, 1987 and references therein), are more likely to show a positive suggestion making behavior (Lipponen et al., 2008), and receive higher ratings for creativity from their managers (Scratchley and Hakstian, 2001). A natural reason for this relationship might be that relatively open workers gained more experience with the merits of using new ways to tackle things, and hence, are also more likely to possess the potential for improving and changing the status quo in organizational settings (George and Zhou, 2001).

In order to check reliability of openness as a measure for creative productivity, I analyze whether the openness score predicts if someone is observed in a creative or non-creative job. I hypothesize that individuals with high scores in openness are more likely to be observed in a creative job. Comparing the scores for openness between the creative and non-creative class yields a significant difference (Wilcoxon rank-sum test,  $p<0.000$ ) confirming that people working in a creative job reveal a higher score in openness than those working in a non-creative job. As this relationship might be driven by other confounders, I run a probit analysis on sorting in creative jobs by controlling for further socio-economic characteristics. Table 7 reveals how the probability to observe an individual in the creative class is predicted by standardized measures of the Big Five personality traits, risk attitude, years of schooling, age, gender and whether the occupation of the subject’s father belongs to the creative sector.

The Big Five openness measure turns out to be a significant predictor of observing someone in a creative job and supports the notion that individuals with relatively high potential to be creative are also more likely to self-select into the creative sector. An openness score which is one standard deviation larger than the mean makes an individual approximately 3 percentage points more likely to work in the creative class.<sup>42</sup> As mentioned earlier, many workers in the

---

<sup>40</sup>Previous evidence also indicates that there exists negative correlations between creative behavior and both conscientiousness and agreeableness, and mostly positive ones for extraversion and neuroticism (Feist, 1998). However, these relationships were observed to be less robust and varied between different dimensions of creativity and different samples.

<sup>41</sup>It might be a concern that personality traits, such as openness, could be endogenous, i.e., it is not only the case that more open people are more likely to sort into the creative sector, but also that working in the creative sector makes people more open. However, evidence by McCrae and Costa (2003) and Soldz and Vaillant (1999) suggests that personality is largely uninfluenced by the environment and remains highly stable over the lifetime.

<sup>42</sup>The Big Five traits conscientiousness and extraversion also show a significant, but negative, relation to the probability of working in the creative class. The link between these two dimensions and creativity have been

creative sector are highly educated, which is confirmed by the significant impact of the years of schooling. Further, there is a positive relation with age while risk aversion and female gender negatively affects the likelihood of being observed in a creative job.

## 5.2 Results

In the following, I present probit models to assess the effect of the creative productivity indicators, risk preferences, gender, and further job and occupational characteristics on the sorting into a job with performance evaluation and performance pay, respectively. According to the experimental results, I expect that years of schooling and the Big Five trait openness will not significantly affect the likelihood that people sort into creative jobs with performance evaluation (pay). In contrast, sorting in the non-creative sector is expected to be positively driven by the years of schooling, as in Dohmen and Falk (2011). Based on previous findings in the literature, I expect that working in a job with performance evaluation (pay) is negatively related to risk aversion and female gender in both sectors. Further, I include experience on the labor market as well as tenure as further productivity indicators in all regressions. First, the likelihood of being affected by performance evaluation (pay) increases in the hierarchical level of an employee which usually requires a certain level of experience or tenure. Second, experience and tenure also increase the expertise and knowledge an employee can be assumed to possess. Expertise has been argued to be an important component for creative output. For instance, expertise can facilitate to imagine or think about complex scientific problems (Amabile, 1996b). Thus, experience and tenure are also expected to be positively related to sorting into jobs with performance evaluation (pay), in creative as well as in non-creative jobs. Table 8 reports marginal effects. In Columns I and IV, the dependent variable “Performance evaluation” takes the value one if the survey respondent indicated being regularly assessed according to performance. In Columns II, III, V and VI, the dependent variable “Performance pay” defines a subgroup of the latter, taking the value one if the respondent indicated being subject to a regular performance evaluation which directly affects her or his gross wage and/or yearly bonus.<sup>43</sup>

For a better illustration of the results, Table 8 reports separate regressions for the creative and non-creative class. A pooled regression with interactions terms for the creative class is not reported here, but shows qualitatively similar results.

First, I consider results for the non-creative sector. Consistent with Dohmen and Falk’s

---

controversially discussed and previous results differ largely depending on the type of the creative task. Therefore, I will focus on openness as the most robust indicator for creative behavior (Feist, 1998; Shalley et al., 2004).

<sup>43</sup>An alternative specification of the dependent variable includes potential long-term salary increases or promotions as “performance pay”. Robustness checks reveal that this alternative variable yields qualitatively similar results.

(2011) previous findings, I also find that the probability of working in a job with regular performance assessment increases significantly with productivity, measured by the years of schooling, experience, and tenure. For instance, one year of schooling raises the probability by 3 percentage points (Column I). Also, risk tolerance positively affects the sorting into jobs with performance evaluation while females are less likely to be observed in such jobs. Additionally, my results indicate that jobs with regular performance evaluation seem to attract people who are relatively less conscientious and more extroverted. If the estimation is restricted to people with performance pay (columns V and VI), coefficients get slightly smaller but stay qualitatively similar. Column VI adds controls for firm size, industry, and occupation dummies. The latter distinguishes between job types (for instance, blue vs. white collar workers) and qualification levels. Dummies range from untrained worker to master craftsmen for blue collar workers and from employee without training/education certificate to employee with managerial responsibilities for white collar workers. Control variables for the public sector and East Germany are also added. Of course, dummies for occupational categories capture a lot of heterogeneity in productivity, that is why coefficients become smaller in size. But the significant impact of all productivity indicators, risk attitude, gender as well as the Big Five trait conscientiousness on sorting remains robust to the inclusion of these controls.

Now I would like to focus on the analysis for the creative class shown in the right columns of Table 8. Here, the years of schooling do not have a significant impact, neither on sorting into a job with performance evaluation nor performance pay. Similarly, the Big Five trait openness has no significant influence on the probability of sorting into a job with performance evaluation and even appears to decrease the likelihood of sorting into performance pay within the creative class. However, the negative relation is not robust to the inclusion of further control variables. These findings support the notion that individuals do not sort into performance pay according to creative ability in jobs or tasks which require creative performance. Experience shows a significant positive impact which is robust to additional controls and emerges for both, sorting into performance evaluation as well as performance pay. For instance, for a worker who has one year more experience above the mean, the probability of working under performance pay increases by 2 percentage points. Tenure only explains a small share of sorting into jobs with performance evaluation whereas the probability of working under performance pay is not affected. In comparison to the non-creative sector, where women are 2.6 percentage points less likely than men to work in a job with regular performance evaluation, the effect is even stronger in the creative sector, where the probability is on average 6.5 percentage points lower for women than for men.

In contrast to the experimental evidence and the findings for the non-creative sector, risk attitude seems to play a minor role for sorting in performance pay within creative jobs. Although coefficient estimates for risk attitude have a similar magnitude for the creative as well as the

non-creative class, the effect is only weakly statistically significant in one of the specifications for the creative class. Finally, I included a measure for reciprocity. In line with Dohmen and Falk (2011), I find a significant impact of reciprocity on sorting into jobs with performance pay for the non-creative sector. However, this relationship is not found for the creative sector. These findings might also be related to the fact that sorting, in practice, has various dimensions which cannot completely be observed in the data. Other preferences such as a strong interest in working in a particular occupation might outweigh others, such as risk attitude or reciprocity. As preferences for certain tasks or jobs are typically not observed in empirical data, it might be an interesting extension for future experimental investigations to focus not only on self-selection into a certain compensation scheme but also into particular task types. As another important extension (Bernard et al., 2014) show that sorting also depends on the availability of alternative contract types. Clearly, in practice, it is usually not a binary choice between, for instance, a fixed pay contract and a tournament based variable pay contract. Other contract types such as piece rate based pay schemes or team-based pay contracts may also be available and affect people's choices.

## 6 Discussion and Conclusion

This paper provides evidence on how agents self-select into a tournament-based compensation vs. a fixed payment scheme in a task which requires creativity.

The results of the experiment reveal that a tournament-based compensation scheme did not attract the most productive agents in terms of creative productivity. In contrast, it attracted agents with higher risk tolerance, higher levels of fun at the task, and more optimistic beliefs about their own rank in the performance distribution. The absence of a significant sorting of the more productive agents into the tournament can be attributed to misjudgments in agents' self-assessments on their relative creative performance. The relative self-assessments of many agents show substantial biases, either in terms of overconfidence or underconfidence. In the psychological literature, different mechanisms are known which bias people's self-assessments. Many of these seem to be particularly relevant for creative tasks, such as the lack of well-defined procedures to generate creative output or the dependence on subjective evaluations which hamper the acquisition of knowledge on one's relative performance (Dunning et al., 2004).

A complementary analysis with a large German survey data set provides additional support to the experimental findings. Here, the Big Five personality trait openness as an established proxy variable for creativity did not predict a positive selection of the more open people into performance pay.

However, I find a large incentive effect of the tournament which is in line with previous evidence

by Bradler et al. (2014) but contradicts predictions from crowding-out theories on the impact of monetary rewards on performance in tasks which are high in intrinsic motivation (Deci et al., 1999). Related to this, the analysis of the data shows no indication that highly intrinsically motivated agents anticipate a crowding-out effect of intrinsic motivation. In contrast, highly intrinsically motivated agents were even significantly more likely to opt for the tournament. Although intrinsic motivation shows a positive correlation with creative performance, it was not large enough to also mediate a significant relationship of creative performance and the selection into the tournament. According to the results, tournament prizes seem to serve as a powerful tool to foster employee performance in creative tasks, but they do not necessarily attract the most creative types.

This result has important implications for the design of incentive contracts and recruiting procedures by organizations. The experimental results have shown that the variable incentive scheme attracted self-confident and risk tolerant agents as well as agents who enjoy performing on the task. If self-confidence, risk tolerance, or pleasure on the job are important attributes for a firm, a variable pay scheme may serve as a screening device. Overoptimism and risk tolerance can be desirable, for instance, to foster aggressive R&D policies (Englmaier, 2010).

However, if a firm cares about attracting people with a great creative potential rather than overconfident agents, careful screening procedures seem to be indispensable, especially in creative jobs. The results may also explain why many organizations in creative and knowledge-intense sectors do not rely on highly incentivized performance pay, such as tournament-based compensation, as a screening device to recruit new talents. In fact, organizations which rely on employees' creativity make use of comprehensive and expensive procedures to identify their preferred candidates. Sophisticated recruiting tools, such as assessment centers, are mostly observed in knowledge-intense professions which constitute a large share of the creative sector, while these screening procedures are less common for professions with simple routine tasks. Assessment centers often select applicants not only with respect to their expertise in a particular field as well as to their social skills, but also according to their creative ability to deal with unusual problems and to improve, for instance, existing products and procedures. Moreover, at art schools, for instance, the unusual uses task is a common component of test procedures used to select eligible candidates. Attracting employees who do not fit to the organization in terms of personal capabilities and characteristics can lead to higher employee turnover, creating additional costs related to recruiting, training, uncertainties in planning, and other processing costs.

Finally, the multi-componential approach of modern creativity research argues that creativity emerges from a complex interaction of individuals' personality traits, their extrinsic and intrinsic motivation, and environmental factors (Amabile, 1996a). Thus, firms that rely on

knowledge-intense and creative output, such as Google or Yahoo<sup>44</sup>, create an inspiring work environment or offer flexible work conditions in order to attract and to motivate talents. These factors can also be thought of as additional sorting dimension individuals might increasingly care about in the future.

The extent to which the likelihood of over- and underestimation of own relative ability improves when employees gain experience on their job and receive repetitive feedback on their relative performance remains an open question for future research. Accurate feedback may enable employees to better assess their relative capabilities and may encourage an efficient sorting of the most capable into performance pay contracts, or more specifically, into tournaments for promotions within a firm. However, as Tversky and Kahneman (1986) argued, basic judgmental biases are unlikely to correct themselves over time. Several reasons have been discussed for the persistence of misjudgments. First, accurate and immediate feedback is rarely available in practice because it is often not easy to attribute outcomes to a particular action or person. Second, the environments and other influential factors often change rapidly and it is difficult to know what the outcome would have been in a different setting. And finally, if feedback is highly subjective, then individuals tend to neglect important information (see, for instance, Dunning et al., 2004). Empirical support for the permanence of biases has been shown by Moore and Cain (2007), where the provision of absolute and relative performance feedback did not reduce biases in relative self-assessments. Even if misjudgments partly correct over time with increasing experience and reliable feedback, wrong self-assessments on own relative ability should be taken into account particularly for young professionals in the creative class when designing incentive schemes to attract creative talents.

---

<sup>44</sup>For instance, James B. Stewart, columnist at the New York Times, reported on Google's philosophy to create an inspiring work environment by providing open kitchens and free-food areas, and by encouraging employees to scribble on walls and to design their own desks like a playground. The article was downloaded on May, 12th via the following link: <http://www.nytimes.com/2013/03/16/business/at-google-a-place-to-work-and-play.html>.

## References

- Acemoglu, D., Autor, D., 2011. Skills, tasks and technologies: Implications for employment and earnings. In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1043–1171.
- Amabile, T., 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45 (2), 357.
- Amabile, T., 1996a. Creativity in context: Update to "The Social Psychology of Creativity.". Westview Press, New York.
- Amabile, T. M., 1996b. Creativity and innovation in organizations. Vol. 5. Harvard Business School Boston.
- Amabile, T. M., 1997. Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review* 40(1), 39–58.
- Amabile, T. M., Hill, K. G., Hennessey, B., M., T. E., 1994. The work preference inventory: Assessing intrinsic and extrinsic motivation orientations. *Journal of Personality and Social Psychology* 66 (5), 950–967.
- 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.
- Avey, J. B., Luthans, F., Hannah, S. T., Sweetman, D., Peterson, C., 2012. Impact of employees' character strengths of wisdom on stress and creative performance. *Human Resource Management Journal* 22 (2), 165–181.
- Bandura, A., 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall, Englewood Cliffs, NJ.
- Banker, R. D., Lee, S.-Y., Potter, G., Srinivasan, D., 2000. An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan. *Journal of Accounting and Economics* 30 (3), 315–350.
- Batey, M., Furnham, A., 2006. Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs* 132 (4), 355–429.
- Bellemare, C., Shearer, B., 2010. Sorting, incentives and risk preferences: Evidence from a field experiment. *Economics Letters* 108 (3), 345–348.
- 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.

- Bernard, M., Dohmen, T., Non, A., Rohde, I., 2014. Performance pay and multidimensional sorting revisited: The role of choice sets, preferences and personality traits, mimeo.
- 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.
- Boschma, R. A., Fritsch, M., 2009. Creative class and regional growth: empirical evidence from seven european countries. *Economic Geography* 85 (4), 391–423.
- Bradler, C., Neckermann, S., Warnke, A. J., 2014. Rewards and performance: A comparison across a creative and a routine task. Mimeo.
- Burt, R. S., 2004. Structural holes and good ideas. *American Journal of Sociology* 110 (2), 349–399.
- Cadsby, C. B., Song, F., Tapon, F., 2007. Sorting and incentive effects of pay for performance: An experimental investigation. *Academy of Management Journal* 50 (2), 387–405.
- Camerer, C., Lovallo, D., 1999. Overconfidence and excess entry: An experimental approach. *American Economic Review*, 306–318.
- Charness, G., Gneezy, U., 2009. Incentives to exercise. *Econometrica* 77 (3), 909–931.
- Charness, G., Grieco, D., 2013. Individual creativity, ex-ante goals and financial incentives, working Paper, Department of Economics, University of California, Santa Barbara.
- Company, M. ., December 2013. Building the creative economy: An interview with Richard Florida.
- Cornelissen, T., Heywood, J. S., Jirjahn, U., 2011. Performance pay, risk attitudes and job satisfaction. *Labour Economics* 18 (2), 229–239.
- Costa, P. T., McCrae, R. R., 1992. Revised neo personality inventory (neo pi-r) and neo five-factor inventory (neo-ffi). Vol. 101. Psychological Assessment Resources Odessa, FL.
- Croson, R., Gneezy, U., 2009. Gender differences in preferences. *Journal of Economic Literature*, 448–474.
- Cruz, S. S., Teixeira, A. A., 2012. Methodological approaches for measuring the creative employment: a critical appraisal with an application to portugal. Tech. rep., Universidade do Porto, Faculdade de Economia do Porto.
- Deci, E. L., 1972. Intrinsic motivation, extrinsic reinforcement and inequity. *Journal of Personality and Social Psychology* 22(1), 113–120.

- Deci, E. L., Koestner, R., Ryan, R. M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6), 627–668.
- Deci, E. L., Koestner, R., Ryan, R. M., 2001. Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research* 71 (1), 1–27.
- Deci, E. L., Ryan, R. M., 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press, New York.
- Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review* 101 (2), 556–590.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G. G., 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9 (3), 522–550.
- Dunning, D., Heath, C., Suls, J. M., 2004. Flawed self-assessment implications for health, education, and the workplace. *Psychological science in the public interest* 5 (3), 69–106.
- Eckel, C. C., Grossman, P. J., 2008. Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results* 1 (7), 1061–73.
- Eisenberger, R., Armeli, S., 1997. Can salient reward increase creative performance without reducing intrinsic creative interest? *Journal of Personality and Social Psychology* 72 (3), 652.
- Eisenberger, R., Aselage, J., 2009. Incremental effects of reward on experienced performance pressure: Positive outcomes for intrinsic interest and creativity. *Journal of Organizational Behavior* 30 (1), 95–117.
- 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., 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., 2010. Managerial optimism and investment choice. *Managerial and Decision Economics* 31 (4), 303–310.

- Eriksson, T., Teyssier, S., Villeval, M.-C., 2009. Self-selection and the efficiency of tournaments. *Economic Inquiry* 47 (3), 530–548.
- Eriksson, T., Villeval, M. C., 2008. Performance-pay, sorting and social motivation. *Journal of Economic Behavior & Organization* 68 (2), 412–421.
- Eysenck, H., 1995. Genius: The natural history of creativity. Cambridge University Press, Cambridge, England.
- Fang, M., Gerhart, B., 2012. Does pay for performance diminish intrinsic interest? *The International Journal of Human Resource Management* 23 (6), 1176–1196.
- Fehr, E., Falk, A., 2002. Psychological foundations of incentives. *European Economic Review* 46 (4-5), 687–724.
- Fehr, E., Fischbacher, U., 2002. Why social preferences matter—The impact of non-selfish motives on competition, cooperation and incentives. *The Economic Journal* 112 (478), C1–C33.
- Feist, G. J., 1998. A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review* 2 (4), 290–309.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10 (2), 171–178.
- Florida, R., 2002. The Rise of the Creative Class. Basic Books New York.
- Florida, R., 2004. The Rise of the Creative Class: And how it's transforming work, leisure, community and everyday life. New York: Basic Books.
- Florida, R., 2006. The flight of the creative class: The new global competition for talent. *Liberal Education* 92 (3), 22–29.
- Frey, B. S., Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys* 15 (5), 589–611.
- George, J. M., Zhou, J., 2001. When openness to experience and conscientiousness are related to creative behavior: an interactional approach. *Journal of Applied Psychology* 86 (3), 513.
- Gneezy, U., Meier, S., Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 191–209.
- Greiner, B., 2004. An online recruitment system for economic experiments. In: *Forschung und wissenschaftliches Rechnen 2003*. Vol. 63. GWDG Bericht, pp. 79–93.

- Grund, C., Sliwka, D., 2010. Evidence on performance pay and risk aversion. *Economics Letters* 106 (1), 8–11.
- Guilford, J., 1967. The nature of human intelligence. New York: McGraw-Hill.
- Guilford, J. P., 1959. Personality. McGraw-Hill, New York.
- Günther, C., Ekinci, N. A., Schwieren, C., Strobel, M., 2010. Women can't jump? – An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization* 75 (3), 395–401.
- Harris, M. M., Schaubroeck, J., 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology* 41 (1), 43–62.
- Hennessey, B., Amabile, T., 1988. The conditions of creativity. In: Sternberg, R. (Ed.), *The nature of creativity*. Vol. 22. Cambridge, UK: Cambridge University Press, pp. 11–38.
- Hyatt, T. A., Taylor, M. H., 2008. The effects of incomplete personal capability knowledge and overconfidence on employment contract selection. *Behavioral Research in Accounting* 20 (2), 37–53.
- 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.
- Kosfeld, M., von Siemens, F. A., 2011. Competition, cooperation, and corporate culture. *The RAND Journal of Economics* 42 (1), 23–43.
- Kruger, J., 1999. Lake wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology* 77 (2), 221.
- Larkin, I., Leider, S., 2012. Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics* 4 (2), 184–214.
- 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.
- Lemieux, T., MacLeod, W. B., Parent, D., 2009. Performance pay and wage inequality. *The Quarterly Journal of Economics* 124 (1), 1–49.
- 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.

- Lipponen, J., Bardi, A., Haapamäki, J., 2008. The interaction between values and organizational identification in predicting suggestion-making at work. *Journal of Occupational and Organizational Psychology* 81 (2), 241–248.
- McCrae, R. R., 1987. Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology* 52 (6), 1258–1265.
- McCrae, R. R., Costa, P. T., 2003. Personality in adulthood: A five-factor theory perspective. Guilford Press.
- Moore, D. A., Cain, D. M., 2007. Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes* 103 (2), 197–213.
- Moore, D. A., Small, D. A., 2007. Error and bias in comparative judgment: on being both better and worse than we think we are. *Journal of Personality and Social Psychology* 92 (6), 972.
- Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics* 122 (3), 1067–1101.
- Prendergast, C., 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1), 7–63.
- Prendergast, C., 2000. What trade-off of risk and incentives? *American Economic Review*, 421–425.
- Runco, M., 1991. Divergent thinking. Ablex Publishing, Norwood, N J.
- Scratchley, L. S., Hakstian, A. R., 2001. The measurement and prediction of managerial creativity. *Creativity Research Journal* 13 (3-4), 367–384.
- 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.
- Simonton, D. K., 2004. Creativity in Science: Chance, logic, genius, and zeitgeist. Cambridge University Press.
- Soldz, S., Vaillant, G. E., 1999. The big five personality traits and the life course: A 45-year longitudinal study. *Journal of Research in Personality* 33 (2), 208–232.
- Stajkovic, A. D., Luthans, F., 1998. Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin* 124 (2), 240–261.
- Sternberg, R. J., 1999. A propulsion model of types of creative contributions. *Review of General Psychology* 3 (2), 83.

- Stiglitz, J. E., 1975. The theory of "screening," education, and the distribution of income. *The American Economic Review* 65 (3), 283–300.
- Torrance, E., 1968. Torrance tests of creative thinking. Personnel Press, Incorporated.
- Torrance, E. P., 1998. *Torrance Tests of Creative Thinking: Norms-technical manual: Figural (streamlined) forms A & B*. Scholastic Testing Service.
- Tversky, A., Kahneman, D., 1986. Judgment under uncertainty: Heuristics and biases. In: Arkes, H., Hammond, K. R. (Eds.), *Judgment and decision making: An interdisciplinary reader*. Cambridge: Cambridge University Press, pp. 38–55.
- Vansteenkiste, M., Deci, E. L., 2003. Competitively contingent rewards and intrinsic motivation: Can losers remain motivated? *Motivation and Emotion* 27 (4), 273–299.
- Waterman, A. S., 2005. When effort is enjoyed: Two studies of intrinsic motivation for personally salient activities. *Motivation and Emotion* 29 (3), 165–188.

## Appendix

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

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

Table 2: Creative Productivity before and after Sorting by Preferred Reward Scheme

	<b>Fixed payment</b>	<b>Tournament</b>	<b>T-test</b>
	Mean (sd)	Mean (sd)	p-value
Productivity Indicator 1 (No reward)	21.891 (10.621)	23.118 (9.758)	0.507
Productivity Indicator 2 (Piece rate)	26.145 (10.997)	27.824 (9.638)	0.370
Productivity in Main Task (Preferred scheme)	23.763 (11.589)	35.853 (11.849)	0.000***
Observations	55	68	

Figure 1: Design of the Experiment

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11
Productivity Indicator 1	Absolute self-assessment	Feedback, Effort Questions	Productivity Indicator 2	Feedback, Effort Questions & Relative Self-Assessment	Sorting Decision	Hypothetical Sorting Decision	Working time for main task	Effort Questions	Risk Preferences (Lottery)	Questionnaire
No payment	What do you think: How did you score in the previous task?	Feedback on score	Piece rate of 10 Taler per point	Individuals can choose between <b>tournament</b> pay scheme (winner gets 1.300 Taler, loser gets nothing) and <b>fixed</b> payment of 400 Taler	Varying the fixed payment alternatives	3 minutes working time under previously chosen contract: tournament or fixed payment	How much fun did you have?	How much effort have you exerted?	Choice between a lottery (400, 0; 0.5)	Demographic Data
3 minutes working time	How much fun did you have?	3 minutes working time	How much fun did you have?	How much effort have you exerted?	How stressed did you feel?	How exhausted did you get?	How stressed did you feel?	How exhausted did you get?	15 safe options: 25, 50, ..., 375	Question on risk attitude (SOEP measure)
	How much effort have you exerted?		How much effort have you exerted?	How stressed did you feel?	What do you think: How many of the 9 other participants scored better than you in the previous task?		One alternative randomly chosen for payoff	One alternative randomly chosen for payoff	WPI measure for intrinsic/extrinsic motivation	Personality Traits (Big Five)
	How stressed did you feel?		How exhausted did you get?				Section for comments			

Figure 2: Productivity of Subjects before Sorting by Preferred Compensation Scheme

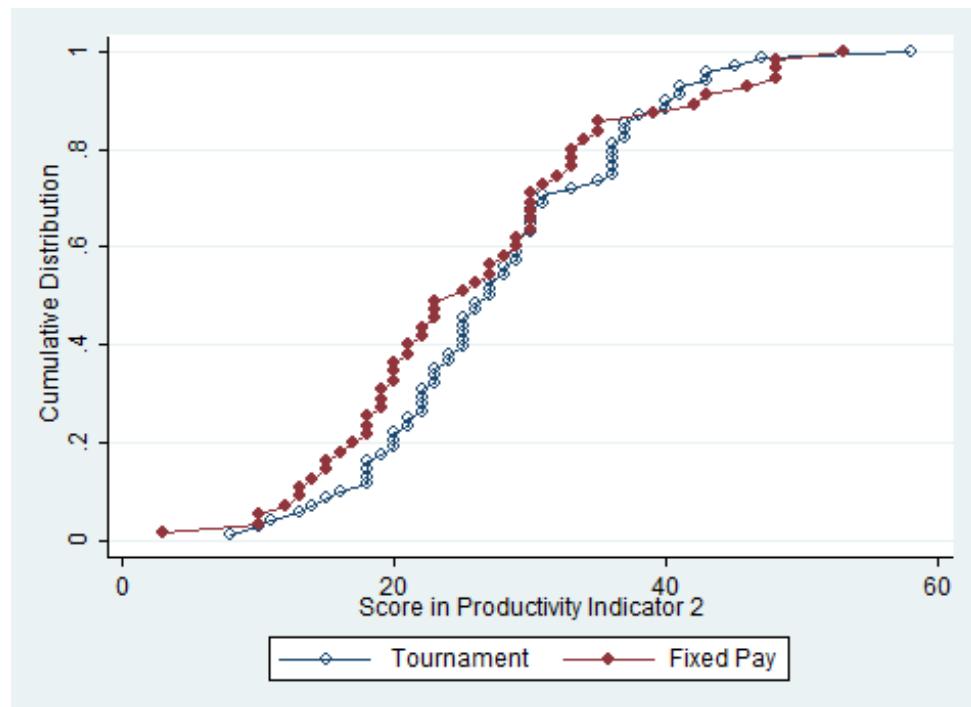


Figure 3: Productivity before Sorting (Productivity Indicator 2) by Score Dimension and Preferred Scheme

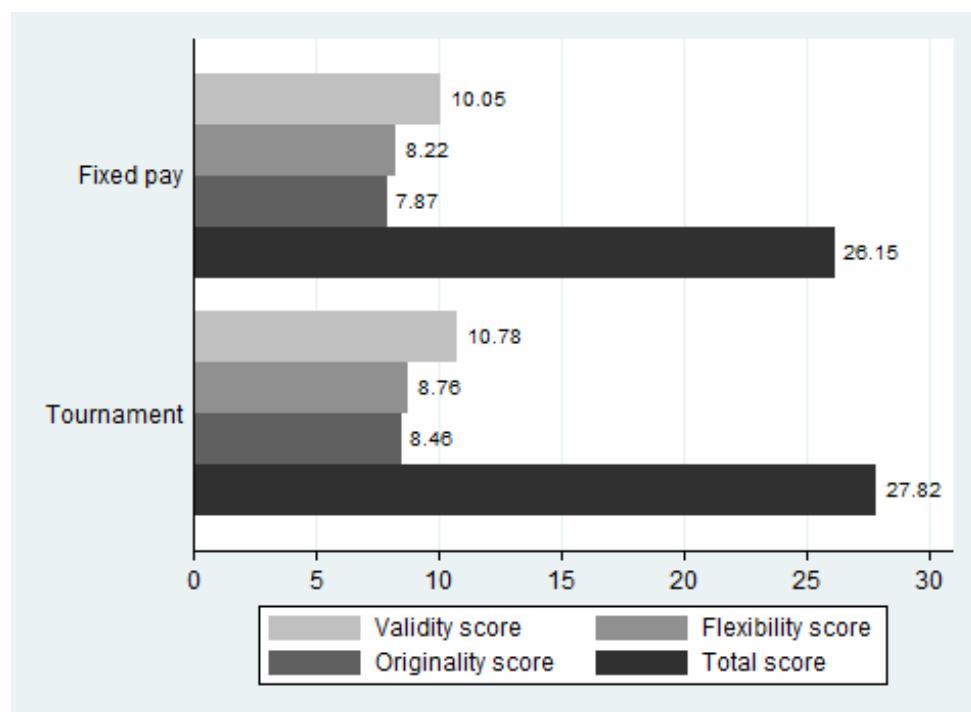


Figure 4: Hypothetical Choices on Sorting by Alternative Fixed Payments

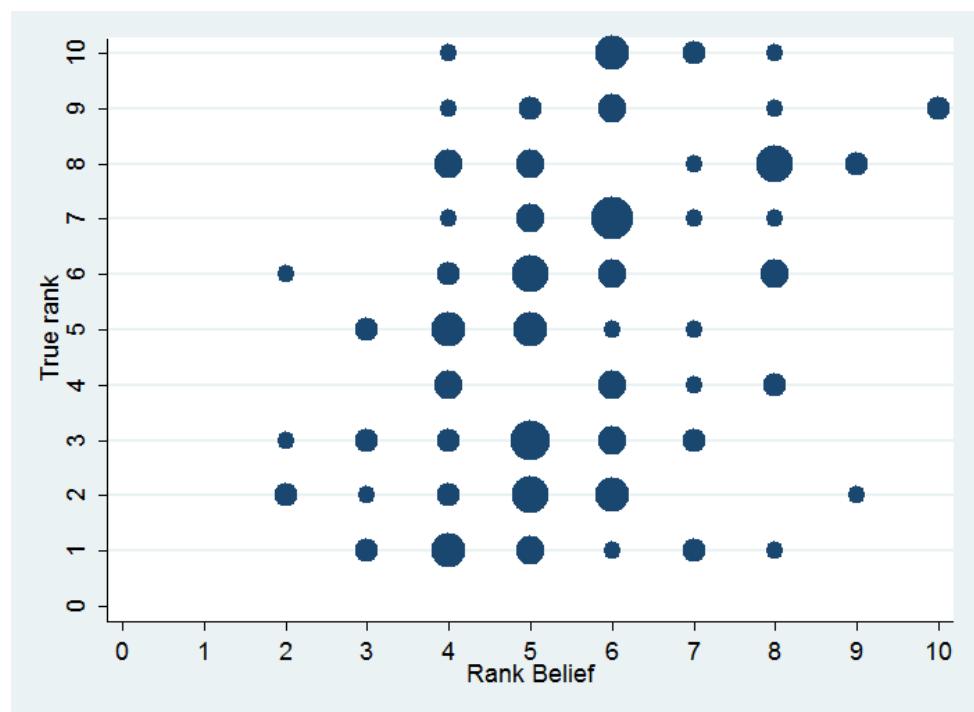


Table 3: Correlations among Independent Variables

	Productivity Indicator 2	Relative self-assessment	Risk Attitude	Female	Fun (Task 2)	Effort (Task 2)	Stress (Task 2)	Exhaustion (Task 2)
Productivity Indicator 2	1							
Relative self-assessment	-0.3541***	1						
Risk Attitude	-0.0329	-0.1564*	1					
Female	0.0863	0.0769	0.0395	1				
Fun (Task 2)	0.3445***	-0.3040***	0.0615	0.2749***	1			
Effort (Task 2)	0.0253	0.0004	-0.0756	0.0981	0.2480***	1		
Stress (Task 2)	-0.1570*	0.2375***	-0.0657	0.1288	0.1002	0.5240***	1	
Exhaustion (Task 2)	-0.0345	0.1028	0.0515	0.1972**	-0.0670	-0.0089	0.3174***	1

*Note:* The Table reports Spearman rank correlations. The variable *relative self-assessment* indicates subjects' belief on the number of better performing participants in the second creative task. Also self-reported measures of perceived fun, effort, stress and exhaustion refer to step 4 of the experiment, that is, the second creative task where subjects worked under the piece rate scheme. Asterisks next to coefficients indicate a significance level as follows: \*\*\*: 1%, \*\*: 5%, \*: 10%.

Figure 5: Subjects Rank Beliefs and True Ranks



*Note:* The size of the spots increase with the observations they contain.

Table 4: Descriptive Statistics on Self-Reported Variables on Fun, Effort, Stress and Exhaustion before and after Sorting

	Before sorting decision			After sorting decision				
	<b>Fixed pay</b>		<b>Tournament</b>	<b>M-W test</b>	<b>Fixed pay</b>		<b>Tournament</b>	<b>M-W test</b>
	Mean	p-value	Mean	p-value	Mean	p-value	Mean	p-value
Fun	3.58	4.34	5.26	0.009***	4.07	4.72	5.97	0.036**
Effort	5.09	0.642	4.44		5.18	5.97	0.000***	
Stress	4.47	3.87	3.00	0.020**	3.33	5.18	0.000***	
Exhaustion	3.18	0.499	3.07		2.96	0.457		
Observations	55	68			55	68		

*Note:* Self-reported measures on fun, effort, stress and exhaustion before sorting refer to the creative task 2, that is, performance under the piece rate (step 5 of the experiment). Asterisks next to coefficients indicate a significance level as follows: \*\*\* 1%, \*\* 5%, \* 10%.

Table 5: Determinants of Sorting in the Tournament (vs. the Fixed Payment Scheme)

Dependent variable	1 if tournament			
	I	II	III	IV
Creative productivity	-0.000 (0.005)	-0.006 (0.005)	-0.000 (0.005)	-0.006 (0.005)
Risk attitude	0.068*** (0.018)	0.074*** (0.020)	0.070*** (0.024)	0.076*** (0.026)
Relative self-assessment	-0.108*** (0.026)	-0.086*** (0.025)	-0.104*** (0.026)	-0.084*** (0.025)
Female	-0.003 (0.062)	-0.077* (0.044)	-0.033 (0.037)	-0.117** (0.049)
Fun		0.089*** (0.034)		0.098*** (0.030)
Effort		0.098** (0.047)		0.110*** (0.040)
Stress		-0.112*** (0.034)		-0.124*** (0.038)
Exhaustion		0.036 (0.029)		0.031 (0.028)
Controls	No	No	Yes	Yes
Observations	123	123	123	123
Sessions	13	13	13	13
Pseudo $R^2$	0.144	0.218	0.156	0.238

*Note:* Probit estimates. Marginal effects (evaluated at the mean of independent variables) are reported. Robust standard errors clustered for sessions are reported in parentheses. The dependent variable is one, if the subject chooses the tournament, and zero if the subject chooses the fixed wage. *Creative productivity* is measured by Productivity Indicator 2 where subjects worked on the unusual task under the piece rate incentive. The variable *relative self-assessment* indicates subjects' belief on the number of better performing participants in the second creative task. That is, the smaller the value of the relative self-assessment, the more productive a subject thinks he or she is relative to others within the session. Self-reported measures on fun, effort, stress and exhaustion refer to the performance under the piece rate before sorting (elicited in step 5 of the experiment). Controls include the Big Five personality measures, age, income and whether a subject has an economic field of study. Asterisks next to coefficients indicate a significance level as follows: \*\*\* 1%, \*\* 5%, \* 10%.

Table 6: Definition of the Creative Class according to ISCO-88 Occupation Codes

Groups of creative people	Occupations and ISCO-88 codes
Creative Core & creative handicraft professions	Physicists, Chemists, and related professionals (211) Mathematicians, statisticians and related professionals (212) Computing professionals (213) Architects, engineers and related professionals (214) Life science professionals (221) Scientific teaching professionals (231-235) Information professionals (243) Social sciences and related professionals (244) Scientific ministerial officers and administrative professionals (247) Musical instrument makers (7312) Jewelry and precious-metal workers (7313) Glass, ceramics and related decorative painters (7324) Handicraft workers in wood, textile, leather and related materials (7331,7332) Tailors, dressmakers and hatters (7433)
Bohemians	Writers and creative performing artists (245) Photographers and image and sound recording equipment operators (3131) Artistic, entertainment and sports associate professionals (347)

*Note:* The definition of the creative class is aligned to Florida (2002) and has been transferred to ISCO-88 occupation codes in a similar way, for instance, by Boschma and Fritsch (2009) or Cruz and Teixeira (2012). Information on the current International Standard Classification of Occupations (ISCO) can be found on <http://www.ilo.org/public/english/bureau/stat/isco/intro.htm>.

Table 7: Sorting into the Creative Class in the German Socio-Economic Panel

Dependent variable	1 if Creative class
Conscientiousness	-0.017*** [0.004]
Extraversion	-0.011*** [0.004]
Agreeableness	-0.001 [0.004]
Openness to experience (std.)	0.025*** [0.004]
Neuroticism	-0.004 [0.004]
Risk attitude	-0.005*** [0.002]
Years of schooling	0.049*** [0.001]
Age (in years)	0.003*** [0.000]
Female	-0.034*** [0.008]
Father in the creative class	0.006 [0.008]
Observations	7,320
Pseudo R-squared	0.291

*Note:* Probit estimates. Marginal effects at the mean reported. Robust standard errors in brackets. The sample consists of merged cross-sectional data from the GSOEP waves 2004, 2008 and 2011. If respondents are observed in more than one of these years, the most recent observation is used. The Big Five have been elicited in the 2005 and 2009 wave of the GSOEP. For the analysis in this paper, I primarily use the information from 2005. If a person was not observed in 2005, I add the information on the Big Five measures of the respective person from the 2009 wave, if available. Asterisks next to coefficients indicate a significance level as follows: \*\*\* 1%, \*\* 5%, \* 10%.

Table 8: Sorting into Performance Pay in the German Socio-Economic Panel

Dependent Variable	Non-Creative Class			Creative Class		
	Performance Evaluation		Performance Pay	Performance Evaluation		Performance Pay
	I	II	III	IV	V	VI
Years of Schooling	0.030*** (0.002)	0.022*** (0.002)	0.009*** (0.002)	0.000 (0.005)	-0.004 (0.004)	0.001 (0.005)
Openness (std.)	0.003 (0.005)	0.001 (0.004)	0.005 (0.004)	-0.019 (0.014)	-0.024** (0.011)	-0.010 (0.011)
Experience full-time (in years)	0.015*** (0.002)	0.012*** (0.002)	0.008*** (0.002)	0.013** (0.005)	0.019*** (0.005)	0.014*** (0.005)
Experience full-time squared	-0.027*** (0.004)	-0.022*** (0.003)	-0.015*** (0.003)	-0.026*** (0.010)	-0.032*** (0.008)	-0.015** (0.008)
Experience part-time (in years)	0.002 (0.003)	-0.002 (0.002)	0.002 (0.002)	0.006 (0.007)	0.005 (0.005)	0.012** (0.006)
Experience part-time squared	-0.007 (0.011)	0.007 (0.008)	-0.006 (0.008)	-0.029 (0.027)	-0.017 (0.024)	-0.042 (0.029)
Tenure	0.006*** (0.001)	0.004*** (0.000)	0.001* (0.000)	0.005*** (0.002)	0.001 (0.001)	0.001 (0.001)
Age (in years)	-0.007*** (0.001)	-0.005*** (0.001)	-0.002** (0.001)	-0.011*** (0.003)	-0.010*** (0.003)	-0.010*** (0.003)
Risk attitude	0.006** (0.002)	0.004** (0.002)	0.005*** (0.002)	-0.004 (0.006)	0.009* (0.005)	0.006 (0.005)
Female	-0.026** (0.012)	-0.035*** (0.009)	-0.028*** (0.010)	-0.065** (0.029)	-0.110*** (0.022)	-0.042* (0.024)
Other Big Five traits						
Conscientiousness (std.)	-0.017*** (0.005)	-0.013*** (0.004)	-0.008* (0.004)	-0.006 (0.013)	0.001 (0.010)	0.014 (0.010)
Extraversion (std.)	0.024*** (0.005)	0.011*** (0.004)	0.004 (0.004)	0.015 (0.013)	-0.007 (0.011)	-0.011 (0.010)
Agreeableness (std.)	-0.006 (0.005)	-0.009** (0.004)	-0.006 (0.004)	-0.008 (0.013)	-0.013 (0.011)	-0.007 (0.010)
Neuroticism (std.)	0.001 (0.005)	-0.011*** (0.004)	-0.003 (0.004)	-0.001 (0.013)	-0.004 (0.011)	0.009 (0.010)
Reciprocity				-0.010** (0.004)		-0.021 (0.013)
Public Sector				-0.019* (0.011)		-0.132*** (0.033)
East Germany				-0.011 (0.009)		0.000 (0.023)
Firm size dummies	No	No	Yes	No	No	Yes
Industry dummies	No	No	Yes	No	No	Yes
Occupation dummies	No	No	Yes	No	No	Yes
Observations	9638	9459	8112	1751	1709	1435
Pseudo $R^2$	0.045	0.063	0.171	0.023	0.063	0.214

Note: Probit estimates. Marginal effects at the mean reported. The sample consists of merged cross-sectional data from the GSOEP waves 2004, 2008 and 2011. If respondents are observed in more than one of these years, the most recent observation is used. The Big Five have been elicited in the 2005 and 2009 wave of the GSOEP. For the analysis in this paper, I primarily use the information from 2005. If a person was not observed in 2005, I add the information on the Big Five measures of the respective person from the 2009 wave, if available. The reciprocity measure is also taken from the 2005 wave. Asterisks next to coefficients indicate a significance level as follows:  
 \*\*\* 1%, \*\* 5%, \* 10%.