



JENA ECONOMIC RESEARCH PAPERS



2014 – 005

Task enjoyment and opportunity costs in the lab – the effect of financial incentives on performance in real effort tasks

by

Katharina M. Eckartz

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich Schiller University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact markus.pasche@uni-jena.de.

Impressum:

Friedrich Schiller University Jena
Carl-Zeiss-Str. 3
D-07743 Jena
www.uni-jena.de

Max Planck Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Task enjoyment and opportunity costs in the lab – the effect of financial incentives on performance in real effort tasks *

Katharina M. Eckartz[†]

February 18, 2014

Abstract

This study is directly motivated by the results of Eckartz et al (2012). Subjects exerted surprisingly high efforts irrespectively of how they were compensated. This paper discusses a number of potential explanations and then it will focus on two of them: first, subjects might exert effort simply because they enjoy working on the tasks. Second, subjects might exert effort because they feel obliged to do so or because they do not have opportunity costs of working. These questions are crucial to better understand the robustness of experimental results and also to be eventually able to transfer the results to the world outside the laboratory. We replicate our earlier results: in the baseline treatment we do not find effects of incentive schemes on the output. Decreasing the attractiveness of the tasks, we also do not observe differences between the incentive schemes. When we introduce, however, a paid outside option, the efforts are higher in the performance-dependent pay treatments than under flat payment. The size of the effect differs between the tasks, the direction is, however, the same.

Keywords: Creativity, Incentives, Real effort task, Experimental methods

JEL Classification: C90, C91, J33

1 Introduction

The experimental economic literature which is studying the effects of financial incentives on subjects' performance reached inconclusive results. In Eckartz, Kirchkamp & Schunk (2012) we compared the effects of three payment schemes in

*I thank Alexander Brehm, Christoph Engel, Juliane Kämmer, Oliver Kirchkamp, Marco Kleine, Arno Riedl and the participants of the IMPRS BBS and the SEEK Workshop on Incentives and Innovations for useful discussions and comments. I thank Claudia Niedlich and Severin Weingarten for assistance and support. I am grateful for financial support by the University of Jena.

[†]Friedrich-Schiller-University Jena, International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World. katharina.eckartz@uni-jena.de

three different real effort tasks. It surprised us that we found, if at all, very small treatment-effects, no matter which task we looked at. Moreover, this outcome is not the result of subjects that are not performing; subjects do exert a significant amount of effort in our experimental tasks regardless of the treatment. While we had competing hypotheses for our target-task (a creative word creation task), we certainly expected treatment effects in the control tasks (solving Ravens' matrices and number adding). The emerging question is why do experimental participants exert substantial effort in real effort tasks when they do not have a financial incentive to do so. Therefore, in this study we want to investigate how sensitive our results are to some of the experimental conditions.

There are a number of potential reasons that might lead the experimental subjects to exerting considerable effort. The first possible explanation is that subjects have a high general or task-specific motivation to perform well. Either because they enjoy working on the tasks and the pleasure of the task offsets their cost of working on it, because they feel challenged and enjoy taking up this challenge, or because of experimental effects. It might be, for example, that the experiment is perceived as an exam situation.¹ The second potential reason is that subjects feel a moral obligation, for example to the experimenter, to exert effort knowing that they will receive a compensation for participating in the experiment. Further explanations are that subjects work on the tasks because they do not have significant opportunity costs of working and they also do not have outside options. The only available outside option in the laboratory is to sit and do nothing, maybe letting your thoughts flow. Consequently, subjects might work only to prevent themselves from being bored. A last explanation, specific to our within-subjects design in Eckartz *et al.* (2012),² is that subjects use the non-contingent pay periods to practice for potential later contingent pay³ periods.

In this study we will investigate two of these possible explanations. First, we aim at changing subjects' work-motivation by making the tasks less fun. Potential ways of achieving this goal are making the tasks more difficult, frustrating or boring. The first two characteristics are linked: making tasks moderately difficult might constitute a challenge, while making them very difficult might make working on them very frustrating. Second, we introduce a paid outside option and thereby aim to investigate jointly the introduction of opportunity cost and the legitimacy of not working. The implemented outside option does not aim at reducing subjects' possibly existing boredom.

With respect to the task characteristics, Camerer & Hogarth (1999) discuss the possible relation between task difficulty and the potential of incentives to influence performance. Camerer & Hogarth argue that this potential is highest for tasks with intermediate difficulty. When the task is very easy, high performance is achievable easily ("floor effect"). When the task is very difficult, even

¹As our subjects were mainly students, our experiment might have created a situation which is similar to an exam. Subjects might be used to exert a lot of effort in such situations.

²The design in Eckartz *et al.* (2012) consisted of 7 stages, in which all incentive schemes were conducted within subjects. The last stage was a self-selection stage. Treatment information were only provided directly before every stage.

³Contingent pay and performance pay will be used interchangeably to subsume linear and tournament payment mechanisms.

though subjects might respond to incentives and increase efforts, it is likely that this increased effort is not resulting in improved results (“ceiling effect”). The importance of skills, and in particular the discrepancy between the needed and the possessed skills, is stressed by Bonner *et al.* (2000). The authors find positive effects of incentives mainly in simple tasks. Furthermore, next to the mere effect of the matching of skills, Brase (2009) argues that also the subjects’ perceived likelihood of increasing output with increased efforts will contribute to the observability of incentive-effects: when subjects rate this likelihood too low, they might give up too early. Ariely *et al.* (2009) analyse the effects of financial incentives in a number of different task types and with different stake-sizes. They find that the effects are, in fact, dependent on the tasks and task characteristics. Relevant for this study is especially their finding that incentives have a positive effect in a key-pressing task but negative effects in a number adding task. Task attractiveness also links to the research by Deci and colleagues on the effects of financial rewards on intrinsic motivation. When the task attractiveness is too high, no or even detrimental effects of financial incentives are to be expected (e.g. Deci *et al.*, 1999). Besides, Bailey & Fessler (2011) looked at both task attractiveness and difficulty in one study. The authors find that pay-for-performance, in their case a piece-rate, was only effective in the unattractive and the non-complex tasks.

When manipulating the task enjoyability, we tried to decrease task enjoyability in the creative task by making the task more difficult, however still solvable. We modified the effort task such that it is easier, but not very attractive.

With respect to the introduction of outside options, different techniques have been applied in the experimental economic literature so far. These techniques varied on the degrees of attractiveness and control: Dickinson (1999) offered subjects the possibility to go home after they completed a minimum number of tasks. He refers to this possibility as “off the job leisure”. He extends the classical labor-leisure-model and considers both the number of hours worked, as well as the effort during the time worked (“on and off the job leisure”). An alternative approach is taken by the studies that introduce time-out buttons in the experiment: Mohnen, Pokorny & Sliwka (2008) included an incentivised *time-out button* in their study. Their focus was on peer pressure in team-work resulting from inequality aversion. In their study students were working independently on a real effort task while their earnings were shared in a 2-person-team. They introduced the time-out button to include an opportunity cost of working. Pushing the time-out button was interpreted as working for the private account. Similarly, Blumkin, Ruffle & Ganun (2010) used an incentivised *stop-button*. Their set-up worked completely with consumption goods (food-vouchers for produced units and drink-vouchers for leisure-time). The focus of their study was to test whether a labor income tax and an equivalent consumption tax lead to an identical labor-leisure allocation. Furthermore, in the study of Gamage, Hayashi & Nakamura (2010) subjects had to make a decision to work for the next 9 minutes and be compensated accordingly or to receive a fixed fee and watch preselected videos instead. The authors’ focus was on the effect of different descriptions of after-tax income on the willingness to work and the amount worked by those who chose to work. A study which aimed at providing attractive outside options to the sub-

Table 1 Treatments

		Paid pause option	
		no	yes
Enjoyability	yes	BL*	Pause
	no	DiffBL	DiffPause

* The baseline equals the original design of Eckartz *et al.* (2012).

jects was conducted by Corgnet, Hernan-Gonzalez & Rassenti (2013). The authors developed a platform on which subjects could easily switch between real effort and real leisure (“on the job leisure”), implemented as surfing on the internet. The authors’ focus was on comparing individual and team incentives as well as on the effect of monitoring. While individual incentives originally outperformed team incentives, the authors found a large positive effect of peer monitoring on efforts.

From this overview, we see that a number of different outside options have been applied in the literature so far. The design in this study will be closest to the one of Mohnen *et al.* (2008) as this design allows subjects to switch between pause and working. Moreover, we can observe how long subjects spend on pause and compensate them accordingly. In addition, this pause button is easy to implement in our experimental setting and allows us to keep a lot of experimental control. This chapter will be organised as follows: we start with introducing the experimental design in Section 2, report the results in Section 3, discuss the findings in Section 4, and will conclude in Section 5. In contrast to the previous paper, here the focus is on the experimental *conditions* under which we observe payment scheme differences.

2 Experiment

In Eckartz *et al.* (2012) we observed that subjects exert substantial effort regardless of the incentive scheme. In this paper we build on the developed design and use a *between-subjects* 2x2 design to test 2 potential explanations for this observation (see Table 1). On the first dimension we vary the enjoyability of the experimental tasks. On the second dimension we introduce a paid pause option. *BL* represents the baseline treatment of this study, *Pause* includes the pause option. Reduced enjoyability will, for simplicity, be labelled as difficulty, abbreviated in the following as *Diff*. The treatments will be described in more detail in Section 2.2. In each of the treatments three different payment schemes are investigated *within-subjects*. Moreover, performance under a self-chosen payment scheme is also investigated. As in the second experimental series of the previous study, two different types of tasks are used: a creative problem solving task and an effort task.⁴ The experiment consists of 8 stages, lasting for 5 minutes each. In contrast to the earlier study, which included only a self-selection state for the creative task, this ex-

⁴The nature of the tasks differs on the enjoyability dimension. The task details will be described in Section 2.3.

periment includes 2 self-selection stages, one for each task. To prevent hedging between the stages, one of the stages is chosen for payment at the end of the experiment randomly by the computer. During the experiment all earnings are displayed as *points* which are converted to Euro at the end of the experiment with a conversion rate of 1 point = 0.04€. Subjects receive the respective experimental instructions on the screen before every stage.

2.1 Hypotheses

The baseline treatment *BL* resembles the set-up of the previous study and, therefore, the first hypothesis relates to the replication of the results of the previous study. The second hypothesis corresponds to the treatment-effect of the first dimension in which the task attractiveness is varied. In the second dimension a paid pause option is introduced. This modification is covered by hypotheses 3a and 3b. The fourth, and final, hypothesis is specific to the self-selection stages in the experiment. All hypotheses are independent of the experimental tasks.

Although classic economic labour supply theory predicts a positive relation between incentives and performance, we did not observe incentive scheme effects in the previous study. As the experimental design matches the essential features of our experiment in Eckartz *et al.* (2012), we expect to replicate the results of the previous study in the *BL* treatment:

Hypothesis 1: replication of results

In *BL* no payment scheme-effects will be observed.

Experimental subjects come to the experiment not only with individual knowledge and skills but also with goals (Camerer & Hogarth, 1999). Next to the goal of earning money, which the experimenter targets at when providing monetary incentives, performance in the experimental tasks will also be influenced by subjects' intrinsic motivation and other individual goals and characteristics. In Eckartz *et al.* (2012) we observed that subjects' performance was, if at all, marginally influenced by the monetary rewards. One potential explanation is that subjects' goals beyond profit maximisation predominate. Therefore, in this experiment we look at two alternative explanations which are captured in hypotheses 2 and 3.

Regarding experimental efforts as costs (as discussed by Camerer & Hogarth, 1999) would imply low efforts under flat payments. It is, however, possible that subjects enjoy working on the experimental tasks and, therefore, the non-monetary benefits arising from working on the task itself outweigh the costs of effort which could lead to the observations that we made in the previous chapter. If this reasoning is relevant, we expect to observe effects of the incentive schemes once the attractiveness of the task is reduced. Therefore, in the experimental manipulation in dimension 1 we try to decrease the task enjoyability by making the tasks more difficult and more boring, respectively:

Hypothesis 2: task enjoyment

In *DiffBL* output under contingent pay will be higher than under flat pay.

The second experimental dimension combines two other explanations. First, it might be that subjects exert substantial effort regardless of the incentive scheme as they feel morally obliged to work.⁵ Second, subjects do not have significant opportunity costs of working in our baseline treatments. Introducing a paid pause option addresses these two points jointly.

We focus on opportunity cost and moral obligations in combination as we expect the latter one to be only a minor reason.⁶ Taking a payoff-maximising perspective, one would expect the effects of a paid outside option to be dependent on the payment scheme. To maximise expected payoffs subjects should go on pause during the whole stage under flat payment. Under performance pay it is payoff-maximising in this experiment to not use the pause option at all.⁷ Therefore, the theoretical benchmarks would be to observe the time on pause to be 5 minutes under flat payment and 0 minutes under performance pay. However, we do not expect these extreme behaviours as subjects are likely to be curious and try out the experimental options they are offered. Also, subjects might trade-off a boring pause against exerting effort in the experimental task. Nevertheless, we expect to observe higher performance under performance pay in *Pause*:

Hypothesis 3a: moral obligation and opportunity costs

In *Pause* output will be higher under performance pay than under flat pay.

The effect of the pause option as motivated in Hypothesis 3a should be independent of the task enjoyability; hence, the same reasoning also applies when in addition to introducing the pause option also the task enjoyability is reduced. If there is an interaction-effect, we anticipate the lower task enjoyability to be reinforcing. Consequently, we expect to observe positive effects of performance pay also in *DiffPause*:

Hypothesis 3b: moral obligation and opportunity costs in less enjoyable tasks

In *DiffPause* output will be higher under performance pay than under flat pay.

If subjects sort according to their preferences or abilities, those subjects who enjoy working on the task or perceive their task-related abilities as being good should

⁵This might be a special case of an experimenter demand effect stemming from one of two sources: subjects may feel obliged to reciprocate and give back something to the experimenter, knowing that they will be compensated for experimental participation. Also by asking subjects in the instructions explicitly to exert effort, our instructions might be loaded. Zizzo (2010) discusses experimenter demand effects in various economic settings. From the studies that Zizzo includes in his paper he concludes that effects of loaded instructions existed but were rather small.

⁶If the pause option is unincentivised, it should only introduce the legitimacy of not working, as subjects see that a possibility to take a break exists.

⁷Under performance pay only subjects with a performance that is smaller than half of the average performance have a higher expected payoff from being on pause than from exerting effort in the experimental task.

choose one of the performance pay schemes and then exert effort in the task.⁸ The other participants should self-select into the flat payment scheme and take a rest. Consequently, in the self-selection stages we expect to see a payment scheme effect in all four treatments:

Hypothesis 4: self-selection

Under self-selection performance will be higher under contingent pay than under flat pay, regardless of the treatment.

2.2 Treatments

In this section we will look at the experimental dimensions and the resulting treatments (Table 1) which were introduced briefly before.

Enjoyability In *DiffBL* and *DiffPause* the characteristics of the experimental tasks were changed with respect to their attractiveness. The aim was to decrease the enjoyability of the tasks by making the tasks more difficult or more boring, respectively. In both tasks we aimed for a modification without expecting the earlier discussed floor and ceiling effects. The exact operationalisation will be explained in Section 2.3 when the experimental tasks are introduced as the implementation differs between the experimental tasks.

Pause option In *Pause* and *DiffPause* a *pause option* like in Mohnen *et al.* (2008) was introduced. The pause option was incentivised, introducing also opportunity costs of working. Subjects could click on the pause-button at any time during the experimental tasks. When the remaining time for the stage was more than 20 seconds, the experimental screen was substituted by a white screen for 20 seconds and subjects received 8 points. When the remaining time was less than 20 seconds, the screen was substituted by a white screen until the end of the stage and subjects received a proportional compensation. Subjects could go on pause as often as they wanted. The opportunity costs were chosen to be about half of the expected payment for that period under performance pay.⁹

The design feature of introducing the outside option in the form of a pause-button with opportunity costs has the advantage of being very clean from an experimental design perspective. All more attractive outside options that were considered in the planning phase of the experiment (e.g. providing magazines, allowing to surf the internet, showing videos, or allowing the subjects to work on their own work) have the disadvantage that a lot of experimental control is lost. Still, such a modification could constitute a point for future research.

⁸The parameters are calibrated such that on average earnings are the same under all three incentive schemes (see Section 2.2). Thus, subjects who are better than the average should choose performance contingent pay to increase their expected payments.

⁹The expected payoff for every stage was 10 Euro, if a subject chose to spend the entire 5 minutes on pause her pause compensation was 5 Euro. Of course, this was only payoff relevant if the particular stage was selected for payment at the end of the experiment.

Table 2 Experimental tasks

Treatment	Creative task	Effort task
BL & Pause	creating words min. word length: 1	adding 5 2-digit-numbers
DiffBL & DiffPause	creating words min. word length: 6	counting 1s in a 5x5 matrix of 0s and 1s

2.3 Tasks

In the *BL*-treatment the used tasks (including the lettersets) were exactly the same as in second experimental series in Eckartz *et al.* (2012). As outlined before, the exact properties of the tasks in this experiment depend on the experimental treatments (see Table 2 for an overview).

Creative task In Eckartz *et al.* (2012) we chose the creative task to fulfil criteria which make it realisable in a computerised laboratory experiment in a within-subjects setting. Amongst others, the task has to be easily repeatable without too high learning-effects. This eliminates insight problems that have been used in other studies.¹⁰ Although tasks that have been used in psychological creativity-research, like for example painting pictures or finding titles for pictures or short-stories¹¹ might remain interesting and challenging over tasks, the quality of the solution is not quickly measurable in an experimental setting.¹² Like in Eckartz *et al.* (2012), we, therefore, decided for a word creation task similar to Scrabble.

Participants were presented with an alphabetically ordered letterset, consisting of 12 letters, e.g. accdeeeeginst. These lettersets were created to be as similar as possible on three dimensions.¹³ Participants were asked to construct as many words as they could within 5 minutes. In the BL and Pause treatments no restrictions on the submitted words were imposed. In DiffBL and DiffPause we introduced a minimum word length of 6 letters to increase the difficulty of the task, this minimum length was based on the results of the pilot experiment.¹⁴

¹⁰Ariely *et al.* (2009), for example, use a task of packing quarters into a box. We also decided against using a sample of these problems as we wanted the tasks to be as similar as possible to each other.

¹¹ For an introductory overview to psychological creativity research, see for example Amabile (1996) or Sternberg (2006).

¹²In tasks like these the quality of the submitted solutions is usually rated by a jury in psychological research. Only recently, Bradler *et al.* (2013) used the “unusual uses” task of the Torrance Test of Creative Thinking (Torrance, 1968) and found a way (based on a pre-test) to relatively quickly rate the submitted solutions such that it was possible to make experimental compensation performance dependent.

¹³We considered the maximum number of points and the total number of words that could theoretically be created as well as the similarity within the solution-sets. Table 24 shows the properties of the used lettersets.

¹⁴We aimed to decrease the task enjoyability; at the same time we wanted to keep the task difficulty still intermediate (as discussed in Camerer & Hogarth, 1999). Based on the pilot experiment, a minimum number of 7 or 8 letters seemed to be too difficult for the subjects, while with a minimum of 6 letters subjects still managed to create a substantial number of solutions.

Table 3 Experimental design

Stage	Letterset / Effort task		Treatment
	BL & Pause	DiffBL & DiffPause	
1	aceehhinrssä	aabeefghllnn	incentive scheme 1
2	adding numbers	counting 1s	
3	aeegglmnru	ceefiiknnstt	incentive scheme 2
4	adding numbers	counting 1s	
5	deehhimnprrt	aeekhknnsstt	incentive scheme 3
6	adding numbers	counting 1s	
7	deegilmnpuw	aeeggiilnnns	self-selection
8	adding numbers	counting 1s	
	questionnaire	questionnaire	

Effort task This task differed between the two treatments. In the BL-treatments, a number adding task, similar to the one used by Niederle & Vesterlund (2007), was used: participants were given five two-digit numbers which they had to add. Their task was to solve as many of these problems as possible within the five minutes of the respective stage. After every task subjects received feedback whether their solution was correct.

In the difficult-treatments we had the choice between making this task more difficult or more boring. We decided to make the effort task more boring and use a pure effort task, namely *counting 1s*. Subjects were presented with a 5x5 matrix consisting of 0s and 1s. The task was to count the number of 1s in that matrix (similar to the task used in Houser, Schunk, Winter & Xiao, 2010).¹⁵ This task was chosen such that everybody can do it and thereby give as little feedback about potentially meaningful skills as possible. Moreover, it is hard to imagine that working on this task is particularly rewarding or fun.¹⁶

Incentive schemes In every treatment of this experiment three different monetary incentive schemes were applied within-subjects: a *flat* fee regime, a *linear* payment regime and a *tournament*. The order of the incentive schemes was altered systematically. The payment schemes were exactly the same as in the previous study. The expected payment in all treatments equaled 10€. This is considerably more than the average hourly wage of a student assistant at the University of Jena. Under a *flat* payment subjects received 10€, no matter how many problems they solved correctly. Under a *linear* payment, subjects in the creative task received for every correct word that they created 1 point for the first letter of the word, 2 points for the second and so on. In the adding-numbers task subjects received 25 points, and in the counting task 6 points for every correctly solved problem.¹⁷ In the *tournament* condition points were allocated like in the linear scheme. However, at the end of the experiment the total number of points acquired in that

¹⁵Ariely *et al.* (2009) found a positive effect in a key-pressing task, which might be comparatively interesting to the *counting 1s* task. They find negative incentive effects in an adding numbers task.

¹⁶In fact, Houser *et al.* (2010, p. 5) designed their task with the goal “to be boring”.

¹⁷Points were converted to Euro at the end of the experiment (see Section 2).

incentive scheme in the particular task were compared to the number of points of three other participants that worked on the task in the same order of incentive schemes. Points acquired from being on Pause did not count for this comparison. If the participant had the highest number of points in this group, she received 25€, otherwise she received 5€.¹⁸

Self-selection The last two stages of the experiment were self-selection stages. Subjects could choose for both tasks separately, which of the previously introduced incentive schemes (flat, linear or tournament) they preferred if that stage was chosen for payment.

Questionnaire In the post-experimental questionnaire socio-demographic variables were collected. Moreover, subjects' risk-preferences were elicited using the risk question that Dohmen *et al.* (2011) tested in the German Socio Economic Panel (SOEP).¹⁹ Besides, we asked for information about subjects' free-time- and language interests. Last, as a manipulation check, subjects were asked how much they enjoyed working on the tasks, how difficult they rate the experimental tasks, and how important it was for them to perform well on the tasks ("task importance"). In the Pause-treatments the questionnaire also included an open-ended question in which subjects were asked to describe how they used the Pause option to possibly get some insights into the motivations underlying subjects' behaviour.

Table 3 gives a detailed summary of the experimental design including the used lettersets. This experimental design allows to analyse the experimental hypotheses about subjects' performance in the experimental tasks. It also allows to observe how much time subjects spend on pause and whether their work intensity differs between treatments or tasks.

Appendix B shows an English translation of the experimental instructions, Appendix B.4 reproduces screenshots of all experimental tasks.

2.4 Conducting the experiment

The experimental sessions were conducted in June and July 2012 in the experimental economic laboratory of the Friedrich-Schiller-University in Jena. In total the experiment was run in 14 sessions, each having between 14 and 18 participants. The first session served as a pilot session.²⁰ In total 228 subjects participated in the experiment of which 195 are considered in the analysis,²¹ 178 of

¹⁸Ties were broken randomly by the computer.

¹⁹This measure consists of a 11-point scale, reaching from 0 (being very risk-averse) to 10 (being very risk-loving).

²⁰The pilot session served to calibrate experimental parameters (in particular how many points each correctly solved task was awarded with) so that the expected payment in all tasks was equal to 10 Euro. Moreover, the pilot session was used to calibrate the minimum word length for the creative task in the difficult treatments.

²¹The pilot session will not be included in the analysis. Moreover, one experimental session will be also excluded from the analysis as we encountered technical problems during that session. Furthermore, one individual participant will be excluded from the analysis as she also encountered a technical problem on her computer.

those were undergraduate students with diverse backgrounds. The mean age of the participants was 23.68. All experimental sessions lasted about one hour. The average payment was 10.76 Euro.

As the experiment included a tournament, a balanced number of men and women was invited.²² Small differences in the respective session-composition are caused by non-show-ups. Overall, the gender composition was balanced: 49% were female. Participants waited in the corridor before the start of the experiment, so they were aware of the balanced gender-composition.

As the creative tasks required very good knowledge of German, in addition to stressing this in the experimental invitation and to conducting a short language test before the start of the experiment,²³ we asked subjects in the post-experimental questionnaire to self-assess their knowledge of German. The participants rated their German skills to be on average 4.9 on a scale from 1 to 5. (see also the left part of Figure 11 in the Appendix).²⁴

Subjects were invited using ORSEE (Greiner, 2004). The experiment was programmed browser-based in PHP in interaction with a MySQL-Database and an Apache-Server. The dictionary we used to check the entered words in the creative task was based on the German isoword-list (Knutzen, 1999).²⁵ Using the Firefox-Plugin *R-Kiosk*, Mozilla Firefox was configured such that the experiment appeared in a full screen mode and subjects could not move back- or forwards in the experiment.

3 Results

This analysis is split by task type. Within the tasks it is guided by the experimental hypotheses developed in Section 2.1. Hypothesis 4, which is specific to behaviour under the self-selected payment scheme, will be dealt with in a separate section.²⁶

To analyse the experimental hypotheses we will rely on mixed effects regressions. This allows to account for the subject-specific heterogeneity.²⁷ We have the choice between estimating a separate regression for every treatment (Equation 1) or estimating a big regression that covers all treatments by including the treatments or treatment characteristics as regressors. We will follow the first approach and analyse every treatment separately as the distributions of the subject-specific random effects ϵ_{subj} differ between the four treatments and as the separate re-

²²Gneezy *et al.* (2003) found that for women's performance in tournaments the gender-composition of the reference group is of relevance. By inviting an equal number of men and women we, therefore, wanted to keep this potential impact on performance constant across sessions.

²³Subjects who passed the language test in earlier experiments did not have to take it again.

²⁴Figure 12 in the Appendix displays subjects' self-reported language interests and hobbies.

²⁵Based on pretest-results this word-list was extended to include more valid words. The words were only included if they were listed in the German online dictionary Duden.de.

²⁶The statistical analysis was done with the statistical software R (R Development Core Team, 2012).

²⁷Due to the experimental design we have multiple observations of the same subject, therefore the observed data is not independent. OLS requires, however, uncorrelated error-terms. Mixed Effects models offer nice possibilities to exploit this experimental design in the regression.

Table 4 Estimation of performance (Equation 1) in the creative task
 Estimation baseline: *flat*

1. Baseline						
	β	σ	t	p value	95% conf	interval
(Intercept)	318	63.3	5.02	0.0000	192	443
linear	2.31	20.9	0.111	0.9121	-39.1	43.8
tournament	-7.8	20.9	-0.374	0.7094	-49.2	33.6
2. Difficult Baseline						
	β	σ	t	p value	95% conf	interval
(Intercept)	218	28.5	7.65	0.0000	161	274
linear	-7.75	18.1	-0.427	0.6704	-43.7	28.3
tournament	-8.45	17.5	-0.482	0.6308	-43.2	26.3
3. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	251	60	4.19	0.0001	132	371
linear	60.4	23.5	2.57	0.0118	13.7	107
tournament	32.7	23.2	1.41	0.1616	-13.4	78.8
4. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	163	26.8	6.07	0.0000	110	216
linear	41.2	20.6	2	0.0480	0.378	82.1
tournament	33.7	20.9	1.62	0.1096	-7.73	75.2

gressions are more intuitive to interpret.²⁸

$$\text{Performance} = \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t} \quad (1)$$

Performance is measured as the number of acquired points. The reference in the estimation is *flat* payment, thus the two contingent pay schemes *linear* payment and *tournament* enter as fixed effects into the regression ($d_{\text{inc.}}$). To account for the within-subjects design, which results in repeated observations of every individual, $\epsilon_{\text{subj.}}$ represents a random effect for every participant. Moreover, a random effect ϵ_{stage} to capture potential differences between the stages is included. The last term, $\epsilon_{\text{subj.,}t}$, is the residual.

Tables 4 and 7 display the regression results of Equation 1 for the creative task and the effort task, respectively. The estimation results are displayed for every treatment separately.²⁹ These tables are the basis for the analysis of hypotheses 1 to 3. The p-values correspond to hypothesis-testing against a two-sided alternative.³⁰ Hypotheses 2 and 3 assume a direction of effect, thus testing against the one-sided alternative would be justified. Nonetheless, we will stick to using the displayed two-sided p-values as these are more conservative.

3.1 Creative task

Hypothesis 1: replication of results To analyse whether the results of the previous study can be replicated, let us look at the first regression in Table 4. In relation to the size of the intercept³¹ ($\beta_0 = 318$) the estimated coefficients are small in magnitude ($\beta_{\text{lin.}} = 2.3$, $\beta_{\text{tourn.}} = -7.8$) and not significantly different from zero. Thus, the result of the previous study can be replicated for the creative task.

Hypothesis 2: task enjoyment This hypothesis is specific to DiffBL (second regression in Table 4). It seems that increasing the task difficulty leads to a general drop in performance independent of the incentive scheme: first, the intercept-size is lower than in BL ($\beta_0 = 217.9$) and second, both performance pay-coefficients are small in magnitude and not significantly different from zero.³² Hence, we do not find evidence in favour of Hypothesis 2.

Hypothesis 3a: moral obligation and opportunity costs The third regression in Table 4 shows the predicted positive impact of performance pay when a pause option is available. Both $\beta_{\text{inc.}}$ have a relatively large magnitude ($\beta_{\text{lin.}} = 60.4$,

²⁸The distributions of $\epsilon_{\text{subj.}}$ are displayed in Figure 6 (for the creative task) and Figure 8 (for the effort task) in the Appendix.

Alternatively, this could be fixed in the big model by interacting $\epsilon_{\text{subj.}}$ with the treatment dummies.

²⁹These regressions are displayed graphically in Figures 7 and 9 in the Appendix.

³⁰The p-values and confidence intervals were bootstrapped using the “mcmcscamp” function in the lme4-package (Bates *et al.*, 2012) in R with 5000 replications.

³¹Here the intercept represents average performance in the BL treatment under flat incentives.

³²Table 14 displays the results of the manipulation check: subjects find the creative task with a minimum number of letters indeed less enjoyable and more difficult than the BL task. Task importance is not influenced.

Table 5 Proportion (in %) of participants who do not use the pause option

1. Creative task						
	Pause			DiffPause		
	β	95% conf. interval		β	95% conf. interval	
Flat	27.91	16.04	42.33	14.29	6.41	25.8
Linear	44.19	30	59.04	26.53	15.57	39.88
Tournament	34.88	21.81	49.71	20.41	10.8	33.03

2. Effort task						
	Pause			DiffPause		
	β	95% conf. interval		β	95% conf. interval	
Flat	39.53	25.84	54.44	20.41	10.8	33.03
Linear	62.79	47.91	76.19	75.51	62.36	86.05
Tournament	39.53	25.84	54.44	73.47	60.12	84.43

β represents the share of participants who do not use the pause option. The confidence intervals were calculated based on a logistic regression in R.

$\beta_{\text{tourn.}} = 32.73$) but only $\beta_{\text{lin.}}$ is significant. Comparing the performance in the different incentive schemes with those in BL, it becomes obvious that the observable effect of the incentive schemes is caused by a drop in performance under flat payments. Thus, we find evidence in favour of Hypothesis 3a, however the effect is significant only for linear incentives. Interestingly, performance under flat payments is still significantly larger than zero.

In the Pause treatments also the time that subjects spent on pause is observable. The time on pause was, as expected, different from the theoretical benchmark: the average time on pause under flat payments was 54 seconds, this is about one fifth of the total time they could spend on pause. Moreover, in all incentive schemes the proportion of subjects who do not spend any time on pause is not negligible (Table 5). This fraction is lowest under flat pay.³³ Introducing a pause option has no impact on work productivity (Table 6). Thus, the observed performance difference results from the longer time on pause that subjects take under flat payments which results in a drop of performance.

Hypothesis 3b: moral obligation and opportunity costs in less enjoyable tasks

The fourth regression in Table 4 tackles Hypothesis 3b. Also here, a positive effect of performance pay is observable. Like in Pause, the difference is caused by the drop in performance under flat payments. As the task characteristics are different from the Pause treatment, the sizes of $\beta_{\text{tourn.}}$ and $\beta_{\text{lin.}}$ in regressions 3 and 4 are not directly comparable, here regression 2 would be the relevant reference for regression 4. Turning to the subjects' productivity: the second regression in Table 6 shows that the work productivity is not significantly different in the performance pay treatments compared to the flat payment.³⁴

³³In Appendix B Table 15 shows that the time on pause was significantly lower under performance pay than under flat pay.

³⁴Looking at how long subjects go on pause, the time on pause is significantly shorter under performance pay (second regression Table 15). Compared to Pause, the time that subjects spend

Table 6 Estimation of Equation 1 for the *productivity* in the creative task, Estimation baseline: *flat*

1. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	0.973	0.182	5.36	0.0000	0.612	1.33
linear	0.119	0.0763	1.56	0.1227	-0.0328	0.271
tournament	0.0434	0.075	0.579	0.5643	-0.106	0.193

2. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	0.685	0.103	6.64	0.0000	0.48	0.89
linear	0.0721	0.0781	0.923	0.3586	-0.0831	0.227
tournament	0.0578	0.0771	0.75	0.4551	-0.0953	0.211

Productivity, measured in points per second, is calculated as the number of points acquired in one stage, divided by the working time of the individual.

3.2 Effort task

Hypothesis 1: replication of results The first regression in Table 7 shows that both $\beta_{\text{inc.}}$ coefficients are, in relation to the size of the intercept ($\beta_0 = 9.04$), small in magnitude ($\beta_{\text{lin.}} = 0.3$, $\beta_{\text{tourn.}} = -0.05$) and no significant effects of performance pay on subjects' performance can be observed. Thus, also for the effort task, the result of Eckartz *et al.* (2012) can be replicated.

Hypothesis 2: task enjoyment The second regression shows the estimation results when the task is the *counting 1s* task. The regression shows that the two performance pay coefficients are not significantly different from zero. Thus, making the task more boring and potentially less challenging³⁵ does not change the impact of the incentive schemes. The different coefficient dimensions, in contrast to regression 1, are due to the different underlying tasks. Note that as also in flat performance is on high levels. Thus, no support for Hypothesis 2 can be found.

Hypothesis 3a: moral obligation and opportunity costs The Pause treatment is designed to deal with this hypothesis. The third regression in Table 7 displays the regression result. Subjects solve more tasks under performance pay ($\beta_{\text{lin.}} = 1$, $\beta_{\text{tourn.}} = 1.74$), however this effect is only significant for tournament pay. Performance under flat incentives is on similar levels like in BL ($\beta_{0 \text{ Pause}} = 8.47$, $\beta_{0 \text{ BL}} = 9.04$). It is significantly higher than zero. Thus, we find evidence in favour of this hypothesis. The effect is, however, not caused by a drop in performance under flat payment.

Looking at the use of the pause option, the time on pause was, as expected, different from the theoretical benchmark: under flat payments subjects spent on

on pause in DiffPause is significantly longer (Table 16).

³⁵Table 18 displays the results of the manipulation check: subjects find the counting 1s number indeed easier than the number adding task. However, task enjoyment and task importance are not different.

Table 7 Estimation of Equation 1: performance in the effort task
 Estimation baseline: *flat*

1. Baseline						
	β	σ	t	p value	95% conf	interval
(Intercept)	9.04	1.62	5.57	0.0000	5.82	12.3
linear	0.325	0.77	0.422	0.6737	-1.2	1.85
tournament	-0.0532	0.779	-0.0684	0.9456	-1.6	1.49

2. Difficult						
	β	σ	t	p value	95% conf	interval
(Intercept)	27.6	1.09	25.2	0.0000	25.4	29.7
linear	0.287	0.745	0.386	0.7006	-1.19	1.77
tournament	0.922	0.732	1.26	0.2110	-0.531	2.38

3. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	8.47	1.34	6.29	0.0000	5.79	11.1
linear	1	0.95	1.05	0.2954	-0.889	2.89
tournament	1.74	0.962	1.81	0.0736	-0.17	3.66

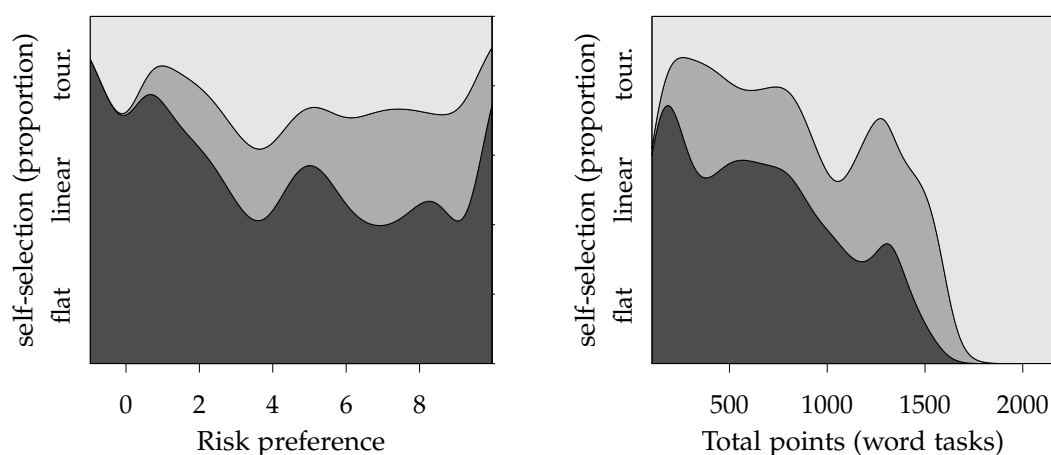
4. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	14	1.8	7.75	0.0000	10.4	17.6
linear	12.3	1.65	7.43	0.0000	9	15.6
tournament	13.1	1.64	8	0.0000	9.87	16.4

Table 8 Estimation of Equation 1 for *productivity* in the effort task
 Estimation baseline: *flat*

1. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	0.039	0.012	3.24	0.0017	0.0151	0.0629
linear	-0.00621	0.0057	-1.09	0.2797	-0.0176	0.00514
tournament	-0.00321	0.00568	-0.566	0.5733	-0.0145	0.00809

2. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	0.0595	0.00823	7.22	0.0000	0.0431	0.0758
linear	0.0308	0.00577	5.34	0.0000	0.0194	0.0423
tournament	0.0335	0.00586	5.71	0.0000	0.0218	0.0451

Productivity, measured in points per second, is calculated as the points acquired in one stage, divided by the working time of the individual.

Figure 1 Self-selection into incentive schemes

average 88 seconds on pause which is a bit less than one third of the stage-length. Also in the effort task the share of participants who do not make use of the pause option is considerable (Table 5). The work productivity is not influenced by the introduction of the pause option (Table 8).³⁶

Hypothesis 3b: moral obligation and opportunity costs in less enjoyable tasks

The regression results for DiffPause are displayed in the fourth regression in Table 7. Note that performance under flat fee is about half of that in DiffBL ($\beta_0 \text{ DiffPause} = 14$, $\beta_0 \text{ DiffBL} = 27.6$). The number of correctly solved tasks is significantly higher with linear and tournament pay.

In this treatment, DiffPause, the work productivity under performance pay is higher than under flat payments (Table 8) and also the time on pause is lower under performance pay.³⁷

3.3 Self-selection

In the last two stages of the experiment subjects have the possibility to select their preferred payment scheme for another round of each of the experimental tasks. Like in the previous section, this analysis is also split by the task type. Subsequently, we will analyse which payment schemes subjects selected and how subjects behaved subsequently.

³⁶Table 19 shows that subjects spent significantly less time on pause under contingent pay than under flat pay.

³⁷ The regression in Table 19 shows that in the counting 1s task subjects spent substantially more time on pause under flat payments. Comparing the time on pause between Pause and DiffPause, Table 20 shows that in the counting 1s task, as compared to the adding numbers task, subjects spent substantially more time on pause under flat payments. Time on pause under linear pay seems to be similar in the two treatments. Under tournament pay, subjects spent slightly less time on pause in DiffPause than in Pause.

Table 9 Self-selection into incentive schemes (in percent): creative task

	Gender	β_{BL}	95% conf	interval	β_{Pause}	95% conf	interval
Flat	f	41.67	23.54	61.48	40.00	20.71	61.71
	m	46.15	28.04	64.99	57.14	36.08	76.56
	all	42.31	29.49	55.86	48.84	34.29	63.52
Linear	f	12.50	3.26	29.29	25.00	9.78	46.25
	m	26.92	12.58	45.61	33.33	15.93	54.58
	all	21.15	11.60	33.48	27.91	16.04	42.33
Tournament	f	45.83	27.09	65.42	35.00	16.83	56.79
	m	26.92	12.58	45.61	9.52	1.65	26.60
	all	36.54	24.33	50.05	23.26	12.40	37.20

	Gender	β_{DiffBL}	95% conf	interval	$\beta_{DiffPause}$	95% conf	interval
Flat	f	72.00	52.84	86.87	50.00	30.77	69.23
	m	50.00	31.47	68.53	36.36	18.58	57.15
	all	60.78	47.10	73.41	46.94	33.39	60.80
Linear	f	20.00	7.68	38.17	16.67	5.50	34.58
	m	19.23	7.36	36.87	31.82	15.12	52.52
	all	19.61	10.35	31.84	22.45	12.35	35.35
Tournament	f	8.00	1.38	22.70	33.33	16.86	53.20
	m	30.77	15.42	49.73	31.82	15.12	52.52
	all	19.61	10.35	31.84	30.61	18.93	44.28

Confidence intervals were calculated based on logistic regressions in R. β represents the share of participants who chose the respective payment scheme.

Table 10 Multinomial logit for incentive selection in stage 7, equation 2

	β	σ	t	p value	95% conf	interval
linear:(intercept)	-2.95	1.09	-2.71	0.0068	-5.08	-0.813
tournament:(intercept)	-3.92	1.1	-3.56	0.0004	-6.07	-1.76
linear:DiffBL	-0.122	0.764	-0.159	0.8736	-1.62	1.38
tournament:DiffBL	0.886	0.735	1.2	0.2283	-0.555	2.33
linear:DiffPause	1.13	0.808	1.4	0.1606	-0.45	2.72
tournament:DiffPause	1.19	0.892	1.34	0.1806	-0.554	2.94
linear:Pause	-0.164	0.718	-0.228	0.8199	-1.57	1.24
tournament:Pause	-1.24	0.949	-1.31	0.1913	-3.1	0.62
linear:points	0.00151	0.00082	1.84	0.0659	-0.000099	0.00311
tournament:points	0.00305	0.000817	3.73	0.0002	0.00145	0.00465
linear:risk	0.206	0.106	1.95	0.0512	-0.00105	0.413
tournament:risk	0.103	0.0994	1.03	0.3009	-0.092	0.298
linear:female	-0.73	0.835	-0.874	0.3824	-2.37	0.908
tournament:female	0.514	0.703	0.731	0.4645	-0.863	1.89
linear:DiffBL:female	0.351	1.12	0.313	0.7540	-1.85	2.55
tournament:DiffBL:female	-2.36	1.14	-2.07	0.0385	-4.59	-0.125
linear:DiffPause:female	-0.335	1.17	-0.286	0.7750	-2.64	1.96
tournament:DiffPause:female	-0.557	1.09	-0.51	0.6102	-2.7	1.59
linear:Pause:female	1.16	1.15	1.01	0.3118	-1.09	3.42
tournament:Pause:female	1.39	1.21	1.15	0.2516	-0.983	3.75

“Flat” is the reference incentive scheme. Effects are shown for the incentive schemes “linear” and “tournament”.

3.3.1 Creative task

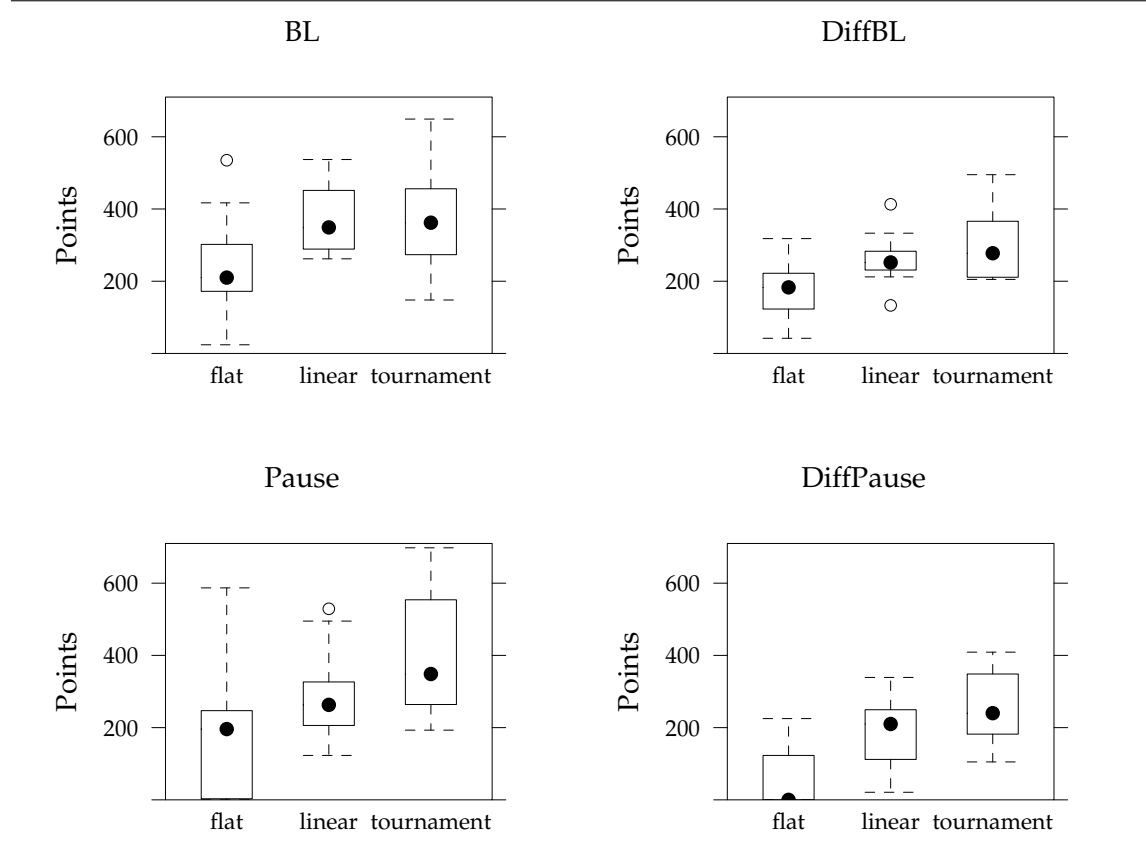
Choices Table 9 shows the self-selection-results, split up by treatment and gender. In all treatments *flat* is the most frequently chosen payment scheme. This is similar to what we observed in the previous study. The share of choices of flat is on similar levels across all treatments and on higher levels in DiffBL. Looking at the genders separately, no additional stable pattern are observable.

Potential determining factors for the self-selection are subjects’ task-specific skills and risk preferences. The left graph in Figure 1 shows that the likelihood of choosing the flat payment scheme seems to decrease with more risk-loving risk preferences. The right graph in the same figure shows clearly that the likelihood of choosing the flat payment scheme decreases with an increasing number of previously acquired points. We take this number of points as a proxy for the task-related ability. To confirm these observations the following multinomial logit model is estimated:

$$\frac{\log \Pr(\text{incentive})}{\log \Pr(\text{flat})} = \beta_{(\text{intercept})} + \beta_{\text{treat.}} \cdot d_{\text{treat.}} + \beta_{\text{points}} \cdot \text{points} + \beta_{\text{risk}} \cdot \text{risk} \quad (2)$$

$$+ \beta_{\text{fem.}} \cdot d_{\text{female}} + \beta_{\text{fem.} \cdot \text{treat.}} \cdot d_{\text{female}} \cdot d_{\text{treat.}}$$

Here flat is the reference incentive scheme, i.e. “incentive” is either linear or tournament. “Treatment” represents the four experimental treatments, where BL represents the base-category in this regression. “Points” is the sum of points obtained in the first three rounds of the creative task. “Risk” is a self-reported risk measure

Figure 2 Performance in stage 7 - creative task

(Dohmen *et al.*, 2011). The higher this measure, the more risk-loving a person.³⁸ Estimation results are reported in Table 10. The first two coefficients show that subjects are less likely to select one of the performance pay schemes than selecting flat pay. Moreover, all β coefficients that are relating to the treatments are not significantly different from zero. The same holds for the effect of gender. Also, except for $\beta_{\text{female-DiffBL}}$ for tournaments, all interaction terms of gender and treatment are not significant. These results confirm the observations from Table 9.

Higher task-related ability makes it more likely for subjects to choose performance pay while the effect is larger for tournament. This effect is significant for both schemes, however only at 10% significance for linear pay. The seemingly low coefficient size is caused by the size that the regressor (total number of points) takes. A similar effect is observable for subjects' risk-preferences. The more risk-loving a subject is, the more likely is the choice of contingent pay, however this effect is only significant for the choice of linear payments. These results confirm our observations from Figure 1.

Performance (Hypothesis 4) Performance in stage 7, as shown in the box-plots in Figure 2, differs between the four treatments even though a general trend is observable across all four treatments: it seems that participants who selected the flat fee obtained fewer points than those who chose performance pay. These re-

³⁸The distribution of the risk measure is shown in the right figure of Figure 11.

Table 11 Self-selection of treatments (in percent): effort task

	Gender	β_{BL}	95% conf	interval	β_{Pause}	95% conf	interval
Flat	f	62.50	42.59	79.87	60.00	38.29	79.29
	m	38.46	21.50	57.60	42.86	23.44	63.92
	all	51.92	38.50	65.16	53.49	38.70	67.87
Linear	f	25.00	10.78	44.29	25.00	9.78	46.25
	m	42.31	24.71	61.35	38.10	19.59	59.35
	all	32.69	20.99	46.06	30.23	17.93	44.83
Tournament	f	12.50	3.26	29.29	15.00	3.96	34.43
	m	19.23	7.36	36.87	19.05	6.34	38.86
	all	15.38	7.34	26.75	16.28	7.34	29.11

	Gender	β_{DiffBL}	95% conf	interval	$\beta_{DiffPause}$	95% conf	interval
Flat	f	48.00	29.34	67.05	37.50	20.13	57.41
	m	26.92	12.58	45.61	31.82	15.12	52.52
	all	37.25	24.85	50.91	34.69	22.39	48.56
Linear	f	32.00	16.11	51.41	54.17	34.58	72.91
	m	42.31	24.71	61.35	59.09	38.39	77.78
	all	37.25	24.85	50.91	57.14	43.19	70.37
Tournament	f	20.00	7.68	38.17	8.33	1.44	23.57
	m	30.77	15.42	49.73	9.09	1.57	25.51
	all	25.49	14.92	38.46	8.16	2.61	17.95

Confidence intervals were calculated based on logistic regressions in R. β represents the share of participants who chose the respective payment scheme.

sults are confirmed by an OLS regression.³⁹ Combined with the observation from the previous paragraph, the experimental results seem to support hypothesis 4 in the creative task. These results are in contrast to the results in stages 1, 3, and 5. In these three stages, when the payment scheme was imposed by the experimenters, we observe effects of the payment schemes only in Pause and DiffPause. Now, when the payment scheme is self-selected, those subjects who self-select into performance pay have a higher output in all treatments.

3.3.2 Effort task

Choices The results of the self-selection for the effort tasks are displayed in Table 11. Unlike in the creative task, *flat* is only the most frequently chosen option in the number adding tasks. Table 12 presents the estimation results of the earlier introduced multinomial logit model (Equation 2) for incentive scheme selection. Here, “points” refers to the performance in the effort task in the earlier stages. In contrast to the creative task some of the treatment-dummies are significant: subjects are more likely to choose tournament in DiffBL than in BL, a similar

³⁹When performance is aggregated on the individual level this leaves one observation per person, therefore we chose to implement a regular OLS regression in place of a mixed effects regression. The regression results are reported in Table 21. β_0 captures performance under flat incentives. A separate regression is run for each treatment.

Table 12 Multinomial logit for incentive selection in stage 8, equation 2

	β	σ	t	p value	95% conf	interval
linear:(intercept)	-3.76	1.09	-3.46	0.0005	-5.89	-1.63
tournament:(intercept)	-5.41	1.38	-3.93	0.0001	-8.11	-2.71
linear:DiffBL	1.17	0.738	1.59	0.1125	-0.275	2.62
tournament:DiffBL	2.2	0.91	2.42	0.0157	0.414	3.98
linear:DiffPause	1.87	0.786	2.38	0.0174	0.329	3.41
tournament:DiffPause	1.43	1.15	1.24	0.2142	-0.827	3.69
linear:Pause	0.159	0.802	0.199	0.8424	-1.41	1.73
tournament:Pause	0.446	0.968	0.46	0.6453	-1.45	2.34
linear:points	0.00355	0.00103	3.44	0.0006	0.00153	0.00558
tournament:points	0.0047	0.0012	3.92	0.0001	0.00235	0.00705
linear:risk	0.226	0.0947	2.39	0.0169	0.0405	0.412
tournament:risk	0.179	0.118	1.52	0.1282	-0.0517	0.41
linear:female	-0.37	0.735	-0.504	0.6145	-1.81	1.07
tournament:female	-0.0193	0.947	-0.0204	0.9837	-1.87	1.84
linear:DiffBL:female	-0.496	1	-0.495	0.6203	-2.46	1.47
tournament:DiffBL:female	-1	1.21	-0.828	0.4074	-3.38	1.37
linear:DiffPause:female	0.0187	0.988	0.0189	0.9849	-1.92	1.95
tournament:DiffPause:female	-0.337	1.48	-0.228	0.8195	-3.23	2.55
linear:Pause:female	-0.449	1.14	-0.395	0.6928	-2.68	1.78
tournament:Pause:female	-0.8	1.42	-0.564	0.5730	-3.58	1.98

“Flat” is the reference incentive scheme. Effects are shown for the incentive schemes “linear” and “tournament”.

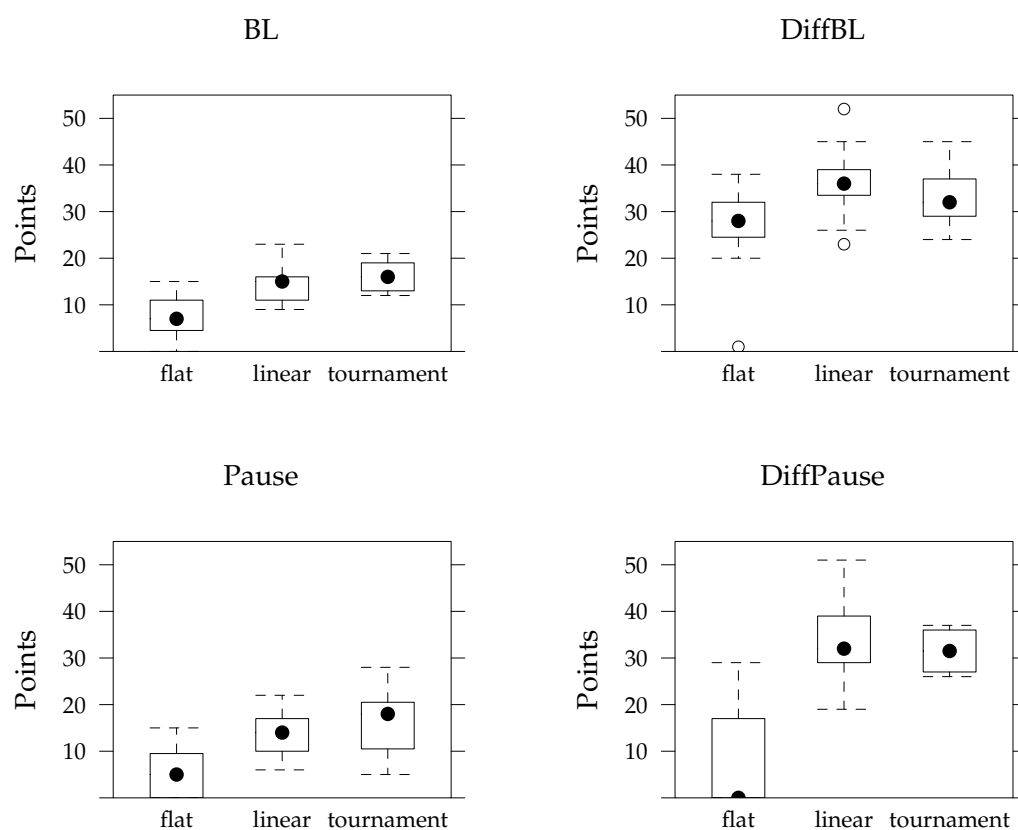
effect is observable in DiffPause when subjects are more likely to choose linear than in BL.

Looking at individual characteristics, gender (β_{female}) has no significant impact on treatment selection, however subjects with a higher task-related ability (β_{points}), i.e. the number of previously acquired points, or higher risk-lovingness (β_{risk}) are more likely to select into performance pay. These effects are highly significant, except for the risk-attitudes in tournament selection.

Performance (Hypothesis 4) Performance in stage 8, as shown in the box-plots in Figure 3, differs between the four treatments but also here a general trend is observable across all four treatments: participants who selected the flat fee obtained fewer points than those who chose performance pay. This observation is supported by an OLS regression.⁴⁰ Combined with the observation that subjects with a higher task-related ability select into performance pay, the experimental results support hypothesis 4 also for the effort task.

These results are in contrast to the results in stages 2, 4, and 6. In these stages, when the payment scheme was imposed by the experimenter, we only observe payment scheme-effects in Pause and DiffPause. Now, when the payment scheme is self-selected, those subjects who self-select into performance pay have a higher output in all treatments.

⁴⁰Estimation results are displayed in Table 22.

Figure 3 Performance stage 8 - effort task

4 Discussion

Since the observed results for the creative and the effort tasks are qualitatively very similar, this discussion will be general. The results of this experimental study can be summed up with two main points.

First of all, subjects' behaviour in the baseline treatments (*BL*), in which the tasks were identical to the earlier study's tasks, does not differ between the three incentive schemes (*flat*, *linear* & *tournament* pay). Thus, we can replicate the results of the previous study. Also, in the difficult baseline (*DiffBL*) conditions, in which the task enjoyability was modified, no effect of the incentive schemes is observed. This result suggests that task attractiveness, task enjoyment or task challenge do not seem to be the major drivers of the experimental results that we observe in the baseline treatment and in the previous study (Eckartz *et al.*, 2012).

Second, with the introduction of a paid pause option (*Pause*) we can observe an effect of the incentive schemes. When subjects have the possibility to make use of the compensated pause option, average performance is higher under performance pay than under flat pay. This effect is driven by a lower performance under flat pay and not by a higher performance under contingent pay.⁴¹ The effect is observable both, when we introduce the pause option alone (*Pause*), as well as when the introduction is combined with a higher task difficulty (*DiffPause*). Interest-

⁴¹There is one exception: in the effort task in the pause-treatment the positive effect of performance pay stems from a higher performance under performance contingent pay.

ingly, within every task subjects' *productivity* is not influenced by the availability of the pause option, with one exception: when subjects had to count 0s and 1s productivity is much lower when a pause option is available. Furthermore, none of the two performance-based incentive schemes seems to outperform the other. Still, it is striking that, even when a paid pause option is available, performance under flat pay is still significantly higher than zero.

As discussed in the introduction to this chapter, with the employed experimental design we cannot distinguish whether the differences in subjects' behaviour result from the legitimacy of not working or the presence of opportunity costs. It is left for future research to disentangle these factors. Some insights into the driving factors might, however, be gotten from the questionnaire responses when subjects were asked to explain how they used the pause option. The most frequently mentioned reasons were physical ones: to take a break, to relax the eyes or because of a lack of concentration. The second most often mentioned groups of reasons are financial and motivational ones: getting a fixed amount of money and not being motivated to continue working. Besides, subjects seem to have used the pause option to receive some compensation for end-of-stage-time and to be compensated while thinking about words in the creative task. No reasons relating directly to legitimacy were given. Yet, taking a break because of physical reasons, when it is officially possible, is possibly be related to legitimacy.

What is striking is that in those conditions in which subjects have not only the chance, but also a financial incentive to take a break (flat in the Pause & DiffPause treatments), we again observe very high effort levels. A substantial fraction of subjects does not use the pause option at all or only for a short time. There are several potential reasons why we can observe this behaviour. First, the reasons that were discussed in the introduction (Section 1) and not targeted by the experimental design, like practicing for later periods or perceiving the experiment as an exam condition, also apply here. In addition, subjects might work as this is what they came to the lab for. By signing up for the experiment they know that this time is dedicated to research and they also know that they will be compensated adequately for their participation. Potentially, the participants perceive signing up for the experiment as a kind of contract between them and the experimenter in which the pause option constitutes a test of loyalty.

Alternatively, it might be that the observed behaviour is a pure subject pool effect. All subjects participated voluntarily and perhaps people who self-selected into the experimental subjects pool are of a very hard-working type, willing to exert effort "in the name of science", no matter how they are compensated. While we cannot exclude this, subject pool effects have been discussed in other experimental settings. For example, a number of authors look at the stability of social preferences in relation to the subject pool.⁴² The studies differ in their conclusions. Some studies find behaviour to be pretty similar in the lab and in the field. Often, however, the observed effects are attenuated in the field. Falk & Fehr (2003) provide a discussion of subject pool effects in the context of labour market experiments and argue that although subject pool effects are important, behaviour does not differ completely between the analysed subject pools. Thus, it

⁴²See, for example, Falk *et al.*, 2011, Exadaktylos *et al.*, 2013, or Levitt & List, 2007.

is rather unlikely that our results are completely due to our specific subject pool.

An effect which is, in the context of our experimental study, probably more relevant, is the impact of subjects' boredom in the laboratory. Even though the pause option is an outside option, it does not provide distraction. Consequently, it might be that subjects work on the experimental tasks as this gives them something to do. By exerting effort also in those conditions where it would be payoff dominant to be on pause during the whole stage, they are indirectly willing to pay for distraction by forgoing the pause compensation. It might be a valuable idea for future research to provide subjects with an attractive outside option. Corgnet *et al.* (2013) for example developed the platform "Virtual Organization" which allows experimental subjects to easily switch between experimental (production) tasks and internet surfing. This tool also allows to track the time of "on the job leisure". Alternatively, at the cost of losing experimental control, providing newspapers or magazines in the laboratory, or allowing the subjects to work on their own work would also reduce the impact of the potentially existing boredom in the laboratory.

The last two stages of the experiment were self-selection stages. In both tasks subjects' performance is, independently of the treatment, higher under self-selected performance pay than under self-selected flat pay. Let us combine this observation with the insight that subjects with a higher number of points in the first stages of the experiment are more likely to self-select into performance pay. It seems that the observed behaviour is the result of sorting by ability (similar to the result obtained by Dohmen & Falk, 2011),⁴³ even though subjects do not receive relative performance feedback. Besides, gender is not significantly related to the choice of payment scheme. Dohmen & Falk have a similar finding, but in their sample risk-attitudes and gender are significantly correlated and therefore risk-attitudes capture the gender effect in their analysis. In our subject pool gender is not significantly correlated with risk attitudes (see Table 23). However, women obtain fewer points in the first stages of the effort tasks. No significant relation between total number of points and gender can be found for the creative task. Hence, our results still seem to be driven by productivity sorting.

5 Conclusion

In Eckartz *et al.* (2012) we examined the effect of three different financial incentive schemes on subjects' performance in three different tasks, one based on creativity, one based on intelligence, and one consisting of adding numbers. After we did not find effects of the different financial incentive schemes in any of the examined tasks, this paper discussed experimental conditions that might have contributed to the result of the earlier study. In this study we relied on an experiment using different real effort tasks: a creative word creation task and two effort tasks (adding numbers and counting 1s). The experiment was designed as a 2x2

⁴³In fact the number adding task is similar in nature to the task that Dohmen & Falk used, namely a math task in which subjects had to multiply numbers.

between-subjects design which varied the incentive schemes within-subjects. On the first dimension we varied the enjoyability of the task, on the second dimension we introduced a paid pause option to introduce opportunity costs and the legitimacy of not working.

First, in the baseline treatment, which replicates the previous study, we, again, observe no effects of the incentive schemes on subjects' performance. Second, when the task enjoyability is reduced, i.e. in the difficult conditions, no effects of the incentive schemes can be observed either. Introducing a paid outside option in the form of a pause-button, however, has an effect: independent of the task we look at, subjects' performance is higher under performance pay than under flat pay. This difference is caused by a lower performance under flat payment when the pause option is available. The size of the difference differs between the tasks, yet the direction is the same across all tasks and levels of difficulty (except for the effort task in the Pause treatment). Interestingly, while under flat pay it would have been payoff-maximising to not work at all and be on pause the whole time, subjects still exert a significant amount of effort and their performance is considerable. In the self-selection stage we observe that under flat pay subjects perform less well than when they self-selected themselves into performance pay.

To conclude, we were aiming at exploring potential reasons why we did not observe effects of financial incentive schemes on subjects' performance in creative and effort tasks. Our results suggests that task attractiveness, task enjoyment or task challenge do not seem to be the main drivers of our earlier observations. However, when subjects have the possibility to make use of an incentivised pause option, their performance stays high under performance pay and decreases under flat pay while their productivity remains unchanged in almost all conditions.

The experimental observations give some directions for potential future research. First of all, with the employed design it cannot be distinguished whether the observed effect in the Pause and diffPause treatments stems from the introduction of opportunity costs or from the fact that also a legitimacy of taking a break is introduced. An experimental treatment which uses a non-compensated pause option could help to shed some light on the driving factor. Observing that subjects exert substantial efforts also under flat pay, and even forego payments in the Pause and difficult Pause treatments when they do not use the pause option as often as possible, indicates that subjects must be driven by something else than pure financial payoff maximisation. Possibly, subjects work to not be bored. Thus, a potential treatment to analyse this point further would be to provide subjects with an attractive outside option to pursue on the job leisure (like e.g. Corgnet *et al.*, 2013).

This study shows that the availability of outside attractive options can be an influential factor and it might be important to give more attention to it when designing economic experiments. In particular, considering that in "real life" outside options are often available, it is important to examine whether experimental results are robust to the availability of attractive outside options. Maybe it is particularly important to keep this in mind in labour economic experiments. This study demonstrated that for creative and effort tasks, independent of the level of difficulty or attractiveness, results are influenced by the availability of an outside

option.

References

- Amabile, T. (1996). *Creativity in context*, Westview press.
- Ariely, D., Gneezy, U., Loewenstein, G. & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies*, 76, pp. 451–469.
- Bailey, C.D. & Fessler, N.J. (2011). The moderating effects of task complexity and task attractiveness on the impact of monetary incentives in repeated tasks. *Journal of Management Accounting Research*, 23(1), pp. 189–210.
- Bates, D., Maechler, M. & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*, R package version 0.999999-0.
- Blumkin, T., Ruffle, B. & Ganun, Y. (2010). *Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods*, Discussion Paper 5145, IZA, Forschungsinstitut zur Zukunft der Arbeit, Bonn.
- 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), pp. 19–64.
- Bradler, C., Neckermann, S. & Warnke, A.J. (2013). *Rewards and performance: A comparison across creative and routine tasks*, mimeo, ZEW Mannheim.
- Brase, G.L. (2009). How different types of participant payments alter task performance. *Judgement and Decision Making*, 4(5), pp. 419–429.
- Camerer, C.F. & Hogarth, R.M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1-3), pp. 7–42.
- Corngnet, B., Hernan-Gonzalez, R. & Rassenti, S. (2013). *Peer pressure and moral hazard in teams: Experimental evidence*, Working paper 13-01, Chapman University, Economic Science Institute.
- 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), pp. 627–668.
- Dickinson, D. (1999). An experimental examination of labor supply and work intensities. *Journal of Labor Economics*, 17(4), pp. 638–670.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. & Wagner, G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*.
- Dohmen, T. & Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *The American Economic Review*, 101(2), pp. 556–590.
- Eckartz, K., Kirchkamp, O. & Schunk, D. (2012). How do incentives affect creativity? *Jena Economic Research Papers*, 6(68).
- Exadaktylos, F., Espín, A.M. & Brañas-Garza, P. (2013). Experimental subjects are not different. *Scientific reports*, 3.

- Falk, A. & Fehr, E. (2003). Why labour market experiments? *Labour Economics*, 10(4), pp. 399–406.
- Falk, A., Meier, S. & Zehnder, C. (2011). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, forthcoming.
- Gamage, D., Hayashi, A. & Nakamura, B. (2010). Experimental evidence of tax framing effects on the work/leisure decision. *Berkeley Olin Program in Law & Economics, Working Paper Series*.
- Gneezy, U., Niederle, M. & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, pp. 1049–1074.
- Greiner, B. (2004). An online recruitment system for economic experiments, in (K. Kremer & V. Macho, eds.), *Forschung und wissenschaftliches Rechnen*, vol. 63, of GWDG Bericht, pp. 79–93, Göttingen: Ges. für Wiss. Datenverarbeitung.
- Houser, D., Schunk, D., Winter, J. & Xiao, E. (2010). *Temptation and commitment in the laboratory*, IEW - Working Papers 488, Institute for Empirical Research in Economics - University of Zurich.
- Knutzen, H. (1999). *hkgerman wordlist*, Technical report, Christian-Albrechts-Universität zu Kiel.
- Levitt, S.D. & List, J.A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), pp. 153–174.
- Mohnen, A., Pokorny, K. & Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4), pp. 693–720.
- Niederle, M. & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3), pp. 1067–1101.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Sternberg, R. (2006). The nature of creativity. *Creativity Research Journal*, 18(1), pp. 87–98.
- Torrance, E.P. (1968). *Torrance tests of creative thinking*, Personnel Press, Incorporated.
- Zizzo, D. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), pp. 75–98.

A Appendix

B Instructions

Subjects first received welcoming instructions in German, followed by task-specific instructions which consisted of a *general part* and a *treatment-specific part*. The following instructions are a translation of the instructions used in the experiment. Unless stated otherwise the instructions apply to all treatments.

Screen 1: Welcoming instructions Welcome to this experiment and thank you for your participation! In total, including reading the instructions, this experiment will last about 1 hour. Please read the following instructions carefully. Should you have any questions, now or at a later point in time, please press the “pause”-key on your keyboard! We will then come to your seat and answer your question in private.

This experiment consists of 8 tasks. For each of these tasks you have 5 minutes. The instructions for the separate tasks will be displayed on your computer-screen before every task. At the end of this experiment one task will be randomly selected for payment.

Important: during the experiment your payoffs will be shown in points. At the end of the experiment the points will be converted to Euro according to the following exchange rate 1 point = 0.04 Euro (i.e. 100 points equal 4 Euro). You will be compensated in cash at the end of the experiment.

[*Only Pause & DiffPause*: During every task you have the possibility to pause the experiment for 20 seconds. For this purpose press the button displayed on your screen. The following rules apply:

- the experimental time continues to run during the time of the break
- you can choose the Pause option as often as you like
- you will be compensated with 8 points for every pause, shorter pause-times at the end of a stage will be compensated proportionally

For example:

- you press the Pause-button at a point in time when you still have more than 20 seconds left to work on your task. Result: your screen will be blank for 20 seconds, you receive 8 points.
- you press the Pause-button at a point in time when you still have 10 seconds left to work on your task. Result: your screen will be blank for 10 seconds, you receive 4 points.]

During the experiment the use of mobile-phones or other tools, as well as the communication with other participants, is not allowed.

B.1 Creative task

General part In this task you are provided with a limited number of letters (a letterset) with which you can create words. The goal of this task is to create as many possible words as possible. *Please enter only one word at a time.* The following rules apply:

- letters may only be used when they are included in the letterset
- within every created word letters may only be used as often as they are included in the letterset
- within each task every word may only be created once
- upper and lower cases are disregarded, only the created words matter
- flections are allowed. If you are presented with the letterset *adenktt* you may use these letters to create, amongst others, *dank, danke, dankt, dankte, danktet*.
 - [Only Diff & DiffPause: the minimal word length is 6 letters. This means in the previous example the words *dank, danke* and *dankt* are too short, the remaining words (*dankte, danktet*) are long enough.]
- Your submissions are compared to a German dictionary. Words that are not included in this dictionary will not be accepted.

The goal is to find as many German words as possible that can be created with the presented letterset. All words included in the German language are in principle allowed.

Not allowed are:

- names of persons or brands
- abbreviations

Please try to find as many different words as possible. Moreover, please try to find long words.

Flat fee If this part is selected for payment, you are rewarded with 250 points, independent of how many words you create.

[Only Pause & DiffPause: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The sum of points for this task will be converted to Euro at the end of the experiment. (Exchange rate 1 point=0.04 Euro)

Tournament If this part is selected for payment, you will receive 1 point for the first letter of the word, 2 points for the second and so on. This means a word consisting of 5 letters would give you 15 points (1+2+3+4+5). Your total number of points will then be compared to the points of three randomly selected participants who work on this task at the same time. The participant with the highest number of points receives 625 points for participating in this experiment. The other three participants receive 125 points each when this task is selected for payment.

[Only Pause & DiffPause: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds. Important: these points will only be added to your points after the comparison with the other participants.]

(Exchange rate 1 point=0.04 Euro)

Linear scheme If this part is selected for payment, you will receive 1 point for the first letter of the word, 2 points for the second and so on. This means a word consisting of 5 letters would give you 15 points ($1+2+3+4+5$).

[*Only Pause & DiffPause*: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The total number of points will be converted to Euro at the end of this experiment. (Exchange rate 1 point=0.04 Euro)

Self-selection If this part is selected for payment, you can decide how you want to be compensated. You may choose between the three possibilities that you have experienced during the previous rounds.

[*Only Pause & DiffPause*: Like in the previous rounds, you will receive in each of the payment schemes, in addition to the payment for the task, also a compensation for your time on pause during this task (8 points for every 20 seconds).]

Please choose which of the three payment schemes you want to be used for your compensation (Exchange rate 1 point=0.04 Euro):

- **Scheme 1:** If this part is selected for payment, you are rewarded with 250 points, independent of how many words you create.
- **Scheme 2:** If this part is selected for payment, you will receive 1 point for the first letter of the word, 2 points for the second and so on. This means a word consisting of 5 letters would give you 15 points ($1+2+3+4+5$). Your total number of points will then be compared to the points of three randomly selected participants in a *previous stage*. You will receive 625 points if your number of points is higher than the number of points of the randomly selected participants. Otherwise, you will receive 125 points.
- **Scheme 3:** If this part is selected for payment, you will receive 1 point for the first letter of the word, 2 points for the second and so on. This means a word consisting of 5 letters would give you 15 points ($1+2+3+4+5$). The total number of points will be converted to Euro at the end of this experiment.

B.2 Effort task (*BL, Pause*): Number adding

General part In this task it is your task to add every time 5 numbers. Please work at your own speed.

Flat Fee If this task is selected for payment, you are rewarded with 250 points, independent of how many tasks you solved correctly.

[*Only Pause*: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The sum of points for this task will be converted to Euro at the end of the experiment. (Exchange rate 1 point=0.04 Euro)

Tournament If this part is selected for payment, you will receive 25 points for every correctly solved task. Your total number of points will then be compared to the points of three randomly selected participants who work on this task at

the same time. The participant with the highest number of points receives 625 points for participating in this experiment. The other three participants receive 125 points each when this task is selected for payment.

[*Only Pause:* In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds. Important: these points will only be added to your points after the comparison with the other participants.]

(Exchange rate 1 point=0.04 Euro)

Linear scheme If this part is selected for payment, you will receive 25 points for every correctly solved task.

[*Only Pause:* In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The total number of points will be converted to Euro at the end of this experiment. (Exchange rate 1 point=0.04 Euro)

Self-selection If this part is selected for payment, you can decide how you want to be compensated. You may choose between the three possibilities that you have experienced during the previous rounds.

[*Only Pause:* Like in the previous rounds, you will receive in each of the payment schemes, in addition to the payment for the task, also a compensation for your time on pause during this task (8 points for every 20 seconds).]

Please choose which of the three payment schemes you want to be used for your compensation.

- **Scheme 1:** If this task is selected for payment, you are rewarded with 250 points, independent of how many tasks you solved correctly. (Exchange rate 1 point=0.04 Euro)
- **Scheme 2:** If this part is selected for payment, you will receive 25 points for every correctly solved task. Your total number of points will then be compared to the points of three randomly selected participants who work on this task at the same time. The participant with the highest number of points receives 625 points for participating in this experiment. The other three participants receive 125 points each when this task is selected for payment. (Exchange rate 1 point=0.04 Euro)
- **Scheme 3:** If this part is selected for payment, you will receive 25 points for every correctly solved task. The total number of points will be converted to Euro at the end of this experiment. (Exchange rate 1 point=0.04 Euro)

B.3 Effort task (*DiffBL, DiffPause*): Counting 1s

General part In this task it is your task to count the number of 1s in a matrix consisting of 0s and 1s.

Flat Fee If this task is selected for payment, you are rewarded with 250 points, independent of how many tasks you solved correctly.

[*Only DiffPause*: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The sum of points for this task will be converted to Euro at the end of the experiment. (Exchange rate 1 point=0.04 Euro)

Tournament If this part is selected for payment, you will receive 6 points for every correctly solved task. Your total number of points will then be compared to the points of three randomly selected participants who work on this task at the same time. The participant with the highest number of points receives 625 points for participating in this experiment. The other three participants receive 125 points each when this task is selected for payment.

[*Only DiffPause*: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds. Important: these points will only be added to your points after the comparison with the other participants.]

(Exchange rate 1 point=0.04 Euro)

Linear scheme If this part is selected for payment, you will receive 6 points for every correctly solved task.

[*Only DiffPause*: In addition, you receive the compensation for your time on pause during this task: 8 points for every 20 seconds.]

The total number of points will be converted to Euro at the end of this experiment. (Exchange rate 1 point=0.04 Euro)

Self-selection If this part is selected for payment, you can decide how you want to be compensated. You may choose between the three possibilities that you have experienced during the previous rounds.

[*Only DiffPause*: Like in the previous rounds, you will receive in each of the payment schemes, in addition to the payment for the task, also a compensation for your time on pause during this task (8 points for every 20 seconds).]

Please choose which of the three payment schemes you want to be used for your compensation.

- **Scheme 1:** If this task is selected for payment, you are rewarded with 250 points, independent of how many tasks you solved correctly. (Exchange rate 1 point=0.04 Euro)
- **Scheme 2:** If this part is selected for payment, you will receive 6 points for every correctly solved task. Your total number of points will then be compared to the points of three randomly selected participants who work on this task at the same time. The participant with the highest number of points receives 625 points for participating in this experiment. The other three participants receive 125 points each when this task is selected for payment. (Exchange rate 1 point=0.04 Euro)
- **Scheme 3:** If this part is selected for payment, you will receive 6 points for every correctly solved task. The total number of points will be converted to Euro at the end of this experiment. (Exchange rate 1 point=0.04 Euro)

B.4 Screenshots

Figure 4 Screenshots creative task

BL & Pause	DiffBL & DiffPause
Sie haben folgende Buchstaben zur Verfügung: aceehhinrssä Mein Wort lautet: Wort bestätigen	Sie haben folgende Buchstaben zur Verfügung: aabeefghllnn Mein Wort lautet: Wort bestätigen Mindest-Wortlänge: 6

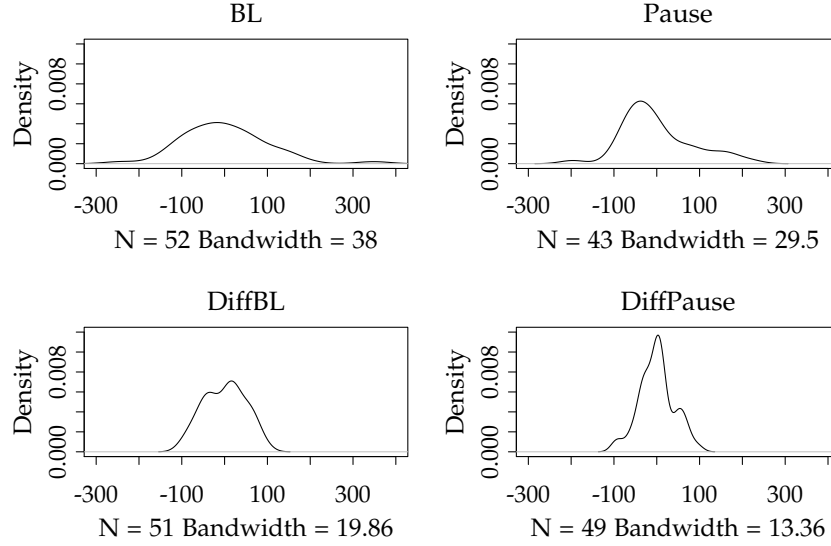
Figure 5 Screenshots effort task

BL & Pause	DiffBL & DiffPause
Bitte addieren Sie die folgenden Zahlen: 21 + 96 + 42 + 90 + 43 Die Lösung lautet: Bestätigen	Bitte zählen Sie die Einsen: 1 0 0 1 0 0 1 1 0 1 0 0 1 1 1 0 0 0 1 0 1 0 0 1 1 Die Lösung lautet: Bestätigen

C Creative task

C.1 Supplementary analysis

Figure 6 Distribution of $\epsilon_{\text{subj.}}$ from Equation 1 for the four treatments



The variances of the subject-specific random effects $\epsilon_{\text{subj.}}$ displayed in Figure 6 are: BL: 9886.55, Pause: 6973.99, DiffBL: 2347.35 and DiffPause: 1657.53.

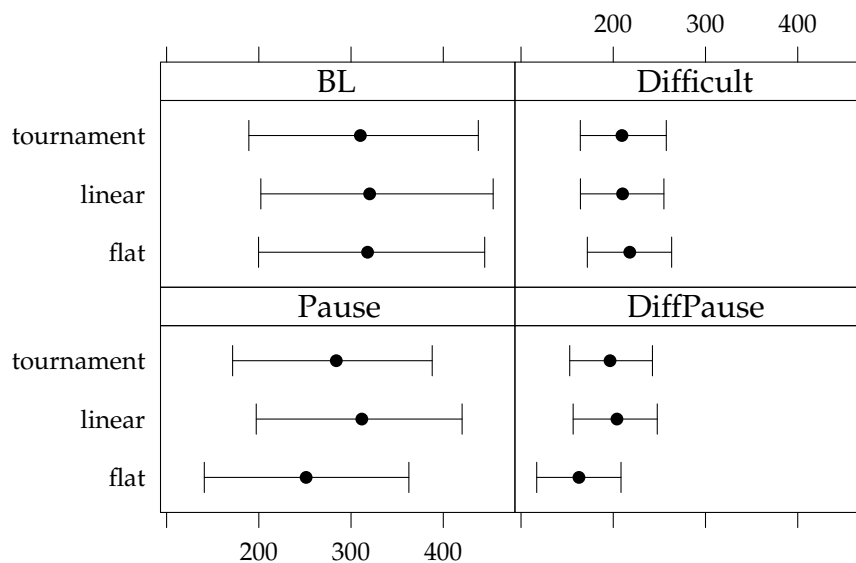
As an alternative to estimating equation 1 separately for each treatment, one could also estimate the following model, the regression results are displayed in Table 13 for the creative task. The experimental dimensions as well as the incentive schemes enter as separate fixed effects as well as in interaction.

$$\begin{aligned}
 \text{Performance} = & \beta_0 + \beta_{\text{diff.}} \cdot d_{\text{diff.}} + \beta_{\text{pause}} \cdot d_{\text{pause}} + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} \\
 & + \sum_{\text{inc.}} \beta_{\text{diff.} \cdot \text{inc.}} \cdot d_{\text{diff.}} \cdot d_{\text{inc.}} + \sum_{\text{inc.}} \beta_{\text{pause} \cdot \text{inc.}} \cdot d_{\text{pause}} \cdot d_{\text{inc.}} \\
 & + \beta_{\text{diff.} \cdot \text{pause}} \cdot d_{\text{diff.}} \cdot d_{\text{pause}} + \sum_{\text{inc.}} \beta_{\text{diff.} \cdot \text{pause} \cdot \text{inc.}} \cdot d_{\text{diff.}} \cdot d_{\text{pause}} \cdot d_{\text{inc.}} \\
 & + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t}
 \end{aligned} \tag{3}$$

Table 13 Estimation of Equation 3: performance in the creative task
Estimation baseline: *flat BL*

	β	σ	t	p value	95% conf	interval
(Intercept)	319	34.5	9.24	0.0000	251	387
diff	-101	22.2	-4.56	0.0000	-145	-57.6
pause	-68.1	23.2	-2.94	0.0035	-114	-22.5
linear	0.888	20.1	0.0442	0.9648	-38.6	40.4
tourn.	-9.7	20.2	-0.48	0.6312	-49.4	30
diff:linear	-6.49	28.9	-0.224	0.8225	-63.4	50.4
diff:tourn.	-0.871	28.5	-0.0305	0.9757	-57	55.2
pause:linear	61.8	29.8	2.07	0.0388	3.19	120
pause:tourn.	41.3	29.9	1.38	0.1673	-17.4	100
diff:pause	12.7	32.3	0.394	0.6938	-50.7	76.1
diff:pause:linear	-14.5	41.9	-0.345	0.7303	-96.8	67.9
diff:pause:tourn.	3.41	41.7	0.0817	0.9350	-78.6	85.4

Figure 7 Performance in the creative task, graphical representation of Table 4



$$\text{Performance} = \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t}$$

Estimates with 95% confidence intervals.

C.2 Task-evaluations

The post-experimental questionnaire contained a number of evaluation-questions ("Evaluation") on how much the subject enjoyed working on the task ("Enjoyment"), how difficult the task was ("Difficulty") and how important a good performance on that task was to the participant ("Importance"). A higher score on the question means a higher task enjoyment, higher task difficulty and a task importance. The following OLS-regression compares the evaluation across the four

experimental treatments. Table 14 displays the regression results for the creative task.

$$\text{Evaluation} = \beta_0 + \sum_{\text{treat.}} \beta_{\text{treat.}} \cdot d_{\text{treat.}} + \epsilon_{\text{subj.}} \quad (4)$$

Table 14 Evaluation of the creative tasks (Eq. 4)

	Enjoyment	Difficulty	Importance
(Intercept)	6.27*** (0.39)	4.65*** (0.27)	6.77*** (0.37)
DiffBL	-1.66*** (0.56)	1.84*** (0.38)	0.11 (0.53)
DiffPause	-0.47 (0.57)	1.20*** (0.38)	-0.14 (0.54)
Pause	0.03 (0.59)	0.09 (0.40)	0.11 (0.56)
R ²	0.06	0.14	0.00
Adj. R ²	0.04	0.13	-0.01
Num. obs.	195	195	195

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

C.3 Time on pause

We will estimate two different mixed effects models to analyse the time that subjects spent on pause. The estimation results of Equation 5 for the creative task are displayed in Table 15. The regression is run for the two pause treatments separately. The estimation baseline is flat pay. The estimation results of Equation 6 are displayed in Table 16. Section D.3 displays the regression results for the effort tasks.

$$\text{Time on pause} = \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t} \quad (5)$$

$$\begin{aligned} \text{Time on pause} = & \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \beta_{\text{DiffPause}} \cdot d_{\text{DiffPause}} \\ & + \sum_{\text{inc.}} \beta_{\text{DiffPause} \cdot \text{inc.}} \cdot d_{\text{DiffPause} \cdot \text{inc.}} + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t} \end{aligned} \quad (6)$$

Table 15 Estimation of Equation 5 for the creative task: time on pause
 Estimation baseline: *flat*

1. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	53.7	17.1	3.13	0.0024	19.6	87.7
linear	-37.6	10.7	-3.5	0.0008	-58.9	-16.2
tournament	-27.7	10.7	-2.59	0.0113	-48.9	-6.42

2. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	85.2	23.4	3.64	0.0004	38.7	132
linear	-50.8	12.7	-4.01	0.0001	-75.9	-25.7
tournament	-50.1	12.6	-3.99	0.0001	-75	-25.2

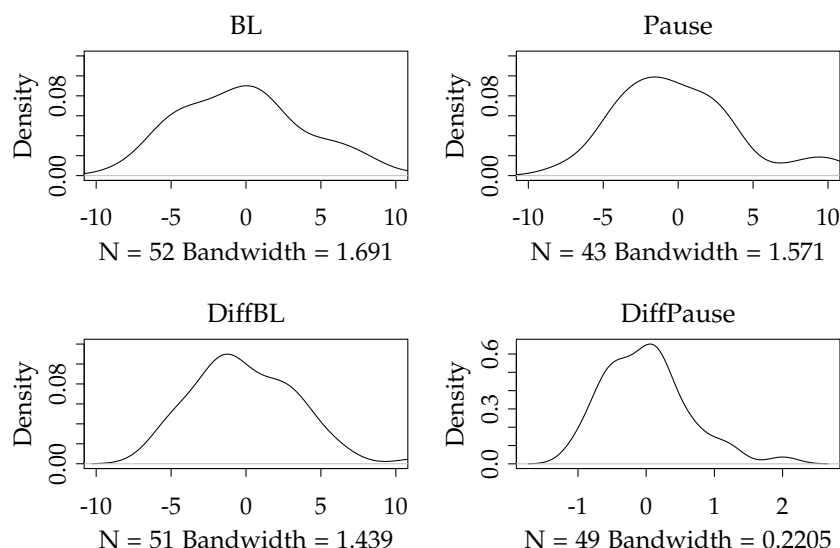
Table 16 Estimation of Equation 6 for the creative task: time on Pause
 Estimation baseline: *Pause-flat*

	β	σ	t	p value	95% conf	interval
(Intercept)	53.7	28.2	1.9	0.0590	-2.07	109
DiffPause	31.5	12.3	2.57	0.0110	7.32	55.7
linear	-37.6	12.4	-3.04	0.0028	-62	-13.2
tournament	-27.7	12.6	-2.2	0.0291	-52.5	-2.85
DiffPause:linear	-13.3	17	-0.782	0.4353	-46.8	20.2
DiffPause:tournament	-22.4	17.2	-1.3	0.1958	-56.4	11.6

D Effort task

D.1 Supplementary analysis

Figure 8 Distribution of $\epsilon_{\text{subj.}}$ from Equation 1 for the four treatments

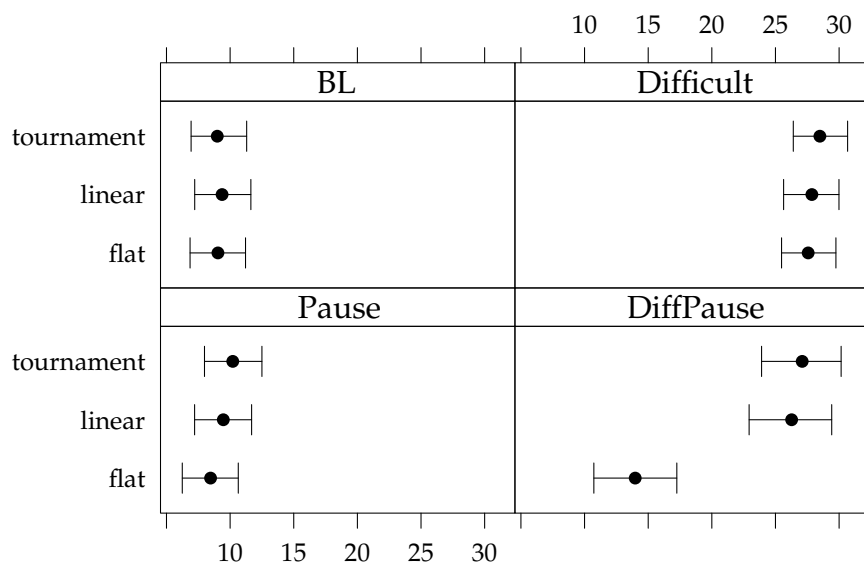


The variances of the subject-specific random effects are: BL: 18.67, Pause: 15.68, DiffBL: 11.68 and DiffPause: 0.36.

Alternatively to estimating Regression 1 separately for every treatment, the following model can be estimated. The experimental dimensions as well as the incentive schemes enter as separate fixed effects as well as in interaction. Table 17 displays the regression results.

Table 17 Estimation of Equation 3: performance in the effort task
Estimation baseline: *flat BL*

	β	σ	t	p value	95% conf	interval
(Intercept)	9.06	1.7	5.33	0.0000	5.72	12.4
diff	18.5	1.16	16	0.0000	16.2	20.8
pause	-0.593	1.2	-0.493	0.6221	-2.95	1.77
linear	0.308	1.08	0.285	0.7758	-1.82	2.43
tournament	-0.0769	1.09	-0.0706	0.9438	-2.22	2.07
diff:linear	-0.0136	1.54	-0.00883	0.9930	-3.04	3.01
diff:tournament	0.998	1.56	0.638	0.5238	-2.08	4.08
pause:linear	0.692	1.6	0.434	0.6646	-2.45	3.83
pause:tournament	1.82	1.62	1.13	0.2606	-1.36	5
diff:pause	-13	1.69	-7.67	0.0000	-16.3	-9.67
diff:pause:linear	11.3	2.22	5.09	0.0000	6.93	15.7
diff:pause:tournament	10.4	2.29	4.54	0.0000	5.88	14.9

Figure 9 Performance in the effort task, graphical representation of Table 7

$$\text{Performance} = \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \epsilon_{\text{stage}} + \epsilon_{\text{subj.}} + \epsilon_{\text{subj.,}t}$$

Estimates with 95% confidence intervals.

D.2 Task-evaluations

Table 18 Evaluation of the effort tasks (Eq. 4)

	Enjoyment	Difficulty	Importance
(Intercept)	4.96*** (0.39)	5.63*** (0.27)	6.77*** (0.37)
DiffBL	0.08 (0.55)	-3.54*** (0.38)	0.11 (0.53)
DiffPause	0.41 (0.56)	-3.66*** (0.39)	-0.14 (0.54)
Pause	-0.38 (0.58)	-0.56 (0.40)	0.11 (0.56)
R ²	0.01	0.43	0.00
Adj. R ²	-0.01	0.42	-0.01
Num. obs.	195	195	195

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

D.3 Time on pause

Table 19 Estimation of Equation 5 for the effort task: time on pause
Estimation baseline: *flat*

1. Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	87.8	19.6	4.47	0.0000	48.8	127
linear	-75.3	15.2	-4.95	0.0000	-106	-45
tournament	-60.3	15.5	-3.9	0.0002	-91.1	-29.5

2. Difficult Pause						
	β	σ	t	p value	95% conf	interval
(Intercept)	123	22.2	5.51	0.0000	78.4	167
linear	-114	12.8	-8.9	0.0000	-140	-88.6
tournament	-114	12.8	-8.85	0.0000	-139	-88.1

Table 20 Estimation of Equation 6 for the effort task: time on Pause
Estimation baseline: *Pause-flat*

	β	σ	t	p value	95% conf	interval
(Intercept)	87.8	22.9	3.83	0.0002	42.6	133
DiffPause	34.8	14.1	2.46	0.0147	6.91	62.6
linear	-75.3	14.3	-5.27	0.0000	-103	-47.1
tournament	-60.3	14.2	-4.23	0.0000	-88.4	-32.2
DiffPause:linear	-38.8	19.8	-1.96	0.0516	-77.8	0.267
DiffPause:tournament	-53.2	19.5	-2.73	0.0070	-91.7	-14.7

E Self-selection

$$\text{Performance (self-selection stage)} = \beta_0 + \sum_{\text{inc.}} \beta_{\text{inc.}} \cdot d_{\text{inc.}} + \epsilon \quad (7)$$

Estimation baseline: *flat*

Table 21 Performance in stage 7 (creative task), estimation of regression 7

	BL	DiffBL	Pause	DiffPause
(Intercept)	239.09*** (24.82)	178.74*** (14.07)	167.29*** (33.57)	62.74*** (18.57)
linear	138.00*** (42.98)	81.46*** (28.50)	115.63** (55.67)	119.08*** (32.64)
tournament	124.54*** (36.45)	120.86*** (28.50)	228.71*** (59.11)	197.13*** (29.55)
R ²	0.24	0.31	0.28	0.50
Adj. R ²	0.21	0.28	0.24	0.48
Num. obs.	52	51	43	49

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 22 Performance in stage 8 (effort task), estimation of regression 7

	BL	DiffBL	Pause	DiffPause
(Intercept)	7.00*** (0.80)	27.37*** (1.66)	5.22*** (1.19)	7.76*** (2.01)
linear	7.41*** (1.29)	8.89*** (2.34)	8.63*** (1.98)	26.16*** (2.55)
tournament	9.13*** (1.67)	6.09** (2.60)	10.93*** (2.47)	23.74*** (4.61)
R ²	0.50	0.24	0.43	0.70
Adj. R ²	0.48	0.20	0.40	0.69
Num. obs.	52	51	43	49

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

F Subject pool

Figure 10 Composition of participants

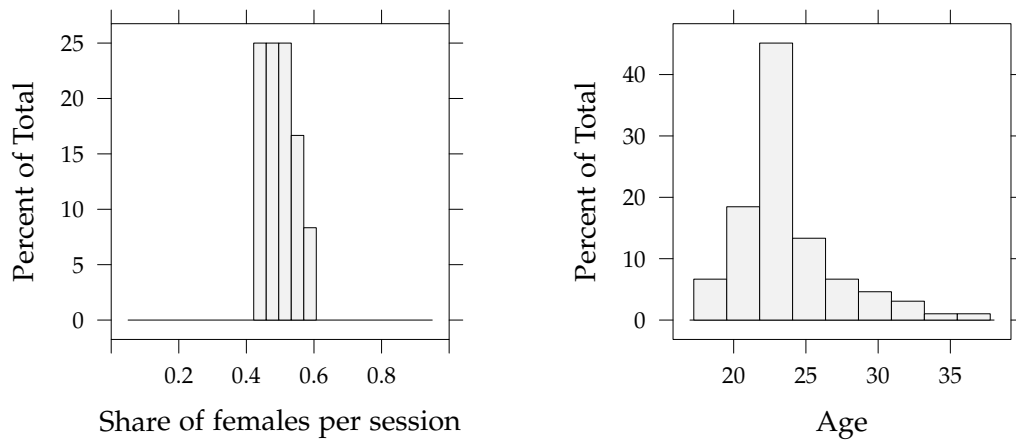
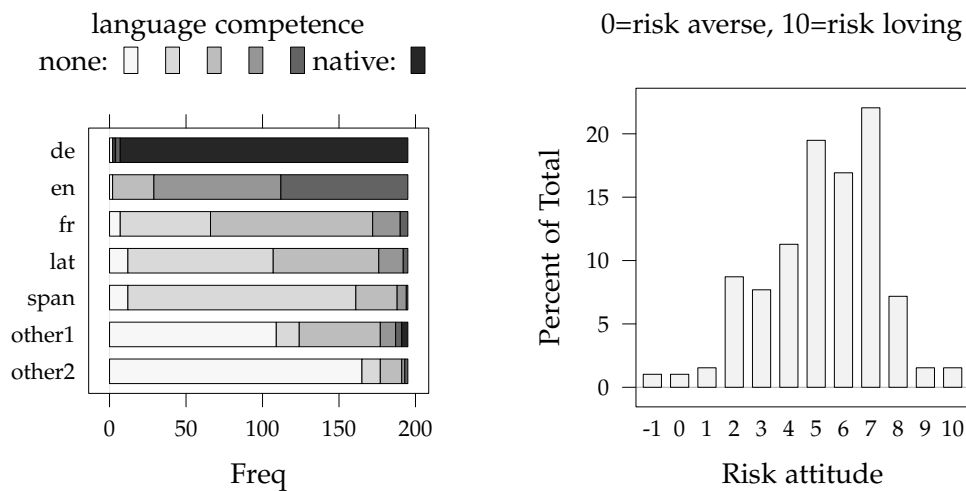
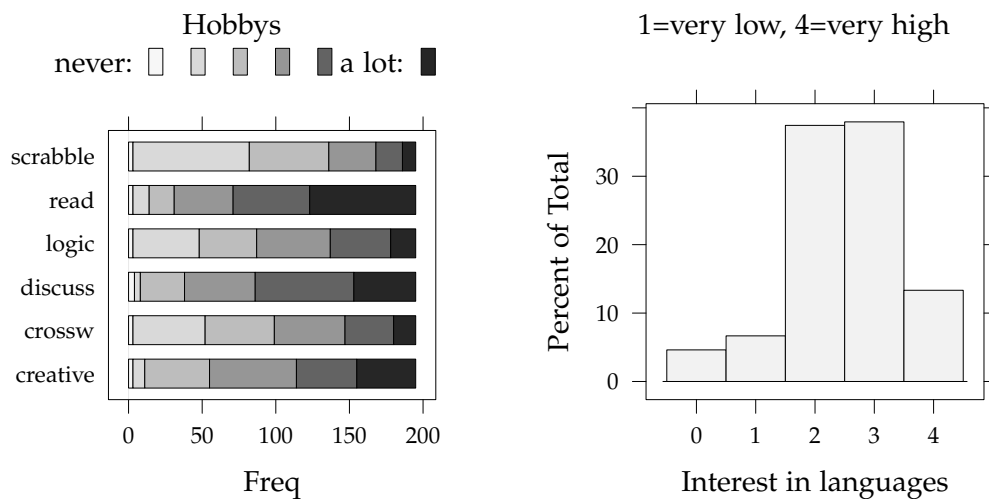


Figure 11 Language competence and risk preferences



Risk attitudes are measured with the SOEP risk-question of Dohmen *et al.* (2011), ranging from 0 to 10, -1 represents NAs.

Figure 12 Hobbies and interest in languages**Table 23** Correlation table (**= $p < .001$, **= $p < .01$, *= $p < .05$)

	female	risk	points crea
female			
risk	-0.05		
points crea	0.05	-0.03	
points effort	-0.16*	-0.05	0.46***

Table 24 Properties of the lettersets

letters	points	words	similarity within
aceehhinrssä	5501	323	0.886879
aeeeggllmnr	5430	323	0.886883
deehhimnprt	5449	321	0.886626
deegilmnpuw	5405	322	0.887139
aabeefghlln	3322	128	0.89621
ceefiiknslt	3313	128	0.895773
aeehknsstt	3352	128	0.89598
aeggiilnns	3320	128	0.896