CBESS Discussion Paper 15-05

Does It Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize

by Emanuela Lezzi*
Piers Fleming**
Daniel John Zizzo***

* School of Economics, University of Insubria
**CBESS and School of Psychology, University of East Anglia
***BENC and Newcastle University Business School, Newcastle University

Abstract
Effort tasks are commonly used to assess individual investment and performance in an experimental setting. Although the tasks used are diverse, they are typically intended to be equivalent as far as they aim to generalize beyond the specific task. We compare an induced value effort task and three real effort tasks in a contest game. Results show that there is no equivalence across tasks in relation to how risk attitude, anxiety and gender predict performance.

JEL classification codes
C72, C90, C91

Keywords
effort tasks, experimental methodology, contests, induced value.

Centre for Behavioural and Experimental Social Science
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/cbess

Does It Matter Which Effort Task You Use? A Comparison of Four Effort Tasks When Agents Compete for a Prize

Emanuela Lezzi*
School of Economics, University of Insubria

Piers Fleming
CBESS and School of Psychology, University of East Anglia

Daniel John Zizzo
BENC and Newcastle University Business School, Newcastle University

Version: April 2015

Abstract

Effort tasks are commonly used to assess individual investment and performance in an experimental setting. Although the tasks used are diverse, they are typically intended to be equivalent as far as they aim to generalize beyond the specific task. We compare an induced value effort task and three real effort tasks in a contest game. Results show that there is no equivalence across tasks in relation to how risk attitude, anxiety and gender predict performance.

Keywords: effort tasks, experimental methodology, contests, induced value.

JEL classification codes: C72, C90, C91.

* Corresponding author. E-mail: emanuela.lezzi@uninsubria.it. Phone: 0332 39 5513; fax: 0332 39 5513; address: via M.te Generoso 71, 21100 Varese, Italy. We wish to thank Robert Sugden, Luigi Mittone, Subhasish Modak Chowdhury, Peter Moffatt and participants at presentation at the University of East Anglia and at a CeDEx -CREED-CBESS Meeting for their helpful comments. We gratefully acknowledge financial support from CBESS, the University of the East Anglia and from an ESRC Interdisciplinary +3 PhD Award.

1. Introduction

It is common for experimental inferences from effort tasks to be drawn just from tasks of a single type. They can be induced value effort tasks in which participants' choice of their level of effort depends on a given cost structure assigned by the experimenter (Smith, 1976), or they can be real effort tasks in relation to which subjects are required to put in non-trivial physical or mental effort. There are a surprising variety of real effort tasks used in economic experiments, such as mathematical questions (Vesterlund and Nierderle, 2007; Eriksson et al., 2008; Dohmen and Falk, 2011; Sutter and Ick-Hannemann 2003, Brüggen and Strobel, 2007), counting numbers (Pokorny, 2008; Abeler et al., 2011; Falk et al., 2006), secretarial tasks (Hennig-Schmidt et al., 2005; Gneezy and List, 2006; Kube et al., 2006; Dickinson, 2001; Konow, 2000; Konow et al., 2009; Falk and Ichino, 2006; Carpenter et al., 2010; Azar, 2009), and computerized tasks (Gill and Prowse, 2012; van Dijk et al., 2001; Bosman et al., 2005; Dickinson and Villeval, 2004; Montmarquette et al., 2004; Sloof and van Praag, 2008). Typically, conclusions that are drawn from these experiments are not qualified by the nature of the task that is employed.[1] Effort tasks are widely used, for example, to predict effort provision in models of individual preferences (Abeler et al., 2011), to predict agents' behaviour in labour markets (Brandts and Charness, 2004), to investigate workers' behaviour (Fehr et al., 1997; Dohmen and Falk, 2011; Falk and Ichino, 2006), to study agents' behaviour in contests (Bull et al., 1987; Harbring and Irlenbusch, 2003; Carpenter et al., 2010), and so on.

The experiment presented in this paper tests whether this implicit equivalence across tasks holds in contests with respect to three specific relationships, one between an emotion (anxiety) and behavior, where the prediction is that anxiety improves effort and therefore performance;[2] another between risk attitude and behavior, where the prediction is that risk attitude decreases effort and therefore performance; and a third in terms of gender differences, where the prediction is that women perform less well than men in contests. Subjects were asked to compete in a two-player all-pay auction to win a prize. They competed by engaging either in an induced value task or in one out of three real effort tasks depending on the treatment. In this competitive setting, we expect a negative relationship between effort, as measured by performance, and risk aversion, and that an anxious emotional state will have a positive effect on performance. The task specificity of our results points to the need to be careful in generalizing from induced value and specific real effort tasks when making experimental inferences.

Section 2 presents the experimental predictions and section 3 the experimental design and procedure. Section 4 contains the results and section 5 the discussion and conclusions.

[1] Examples include Dohmen and Falk (2011), Charness and Villeval (2009), and Falk et al. (2006).

[2] Predictions are discussed in section 2. Note that one of our tasks is an induced value task, and it would be more usual to speak of investment rather than performance in this setting. For simplicity, however, we simply refer to performance to all tasks in this paper.

2. Experimental predictions

In order to test for any equivalence between four effort tasks chosen within the most popular tasks in experimental settings, we ran an economic experiment which consists of four identical treatments except for the effort task used. The tasks were three real effort tasks and an induced value effort task. The real effort tasks were: the slider task (Gill and Prowse, 2012), the mathematical task (Vesterlund and Nierderle, 2007), and (a variation of) the counting task (Abeler et al., 2011).

In this section we first establish the three benchmark predictions, in relation to anxiety, risk attitude and gender, and we then explain why simply inferring conclusions from a specific task may not apply equally across all tasks.

Benchmark predictions: anxiety. In our experiment subjects are asked to expend costly effort to compete in a contest for a prize. Subjects' decision of the level of exerted effort may depend on the arousal of anxiety that occurs in a competitive setting. Our benchmark prediction is that anxiety improves subjects' performance (*Hypothesis* 1). This hypothesis is supported by the Processing Efficiency Theory (Eysenck and Calvo, 1992), according to which anxiety has a prevailing motivational function that induces the subject to increase effort and therefore to boost performance in order to avoid the likely negative implications of poor performance. This seems reasonable for the at most moderate levels of anxiety subjects are likely to experience in our laboratory settings.

Benchmark predictions: risk attitude. Millner and Pratt (1991) carried out an experiment that investigates the possibility that risk attitude may explain rent over-dissipation, as measured by the level of expenditure in lottery tickets. They grouped together more risk averse participants and they compared effort between groups. The more risk averse group exerted lower effort in aggregate than the less risk averse group. Even though theory predictions can in principle be ambiguous (Konrad and Schlesinger, 1997), the experimental results fit with the standard theoretical prediction that higher risk aversion leads to lower effort.[3] We therefore expect that more risk averse individuals will exert less effort than less risk averse individuals (*Hypothesis* 2).

Benchmark predictions: gender. The existing experimental evidence shows that women do not perform as well as men under contest incentives (Gneezy et al., 2003; Gneezy and Rustichini, 2004), and this is therefore our prediction (*Hypothesis* 3).

Across tasks differences. The above benchmark predictions are not task-specific. However, there are good reasons to believe that the task employed may make a difference in drawing conclusions from effort tasks (*Hypothesis* 4). This is, perhaps, more straightforward to see in comparing induced value tasks with real effort tasks, for a key part of the motivation for real effort experiments has always been to accept a loss of experimental control relative to induced value experiments in exchange for having tasks that entail actual effort and are therefore closer to natural settings in this key respect. Our experiment uses contests and the

[3] See Hilman and Katz (1984), Skaperdas and Gan (1997) and Treich (2010).

tight connection between theory and experiments in contests has meant a frequent use of the induced value methodology (see for example, Bull et al., 1987; Schotter and Weigelt, 1992; Zizzo, 2002; Harbring and Irlenbusch, 2003, 2005; Harbring and Lünser, 2008; Breitmoser et al., 2010; Sheremeta and Wu, 2012).[4]

Although real effort tasks promise better external validity at a cost of some interpretability,[5] there are good reasons to expect that there may be different responses depending on the nature of the effort tasks. Some of them could require cognitive skills or physical effort, which might reasonably affect performance, whereas others could require only 'procedural knowledge'[6]. Real clerical tasks may differ in task learning, required skills and intrinsic motivation. In addition, the implementation of real effort tasks might not always be carried out controlling for the heterogeneity in the distribution of the abilities.

3. The Experiment: Design and Procedure

The experiment was conducted at the University of East Anglia in Norwich, U.K. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). 206 subjects (46% male and 54% female) took part in the 24 sessions. Each session lasted approximately one hour.[7] Sessions consisted of three stages: 1) first, subjects filled in an anxiety questionnaire; 2) they then participated in 10 rounds of a 2-player contest; and 3) finally they completed a risk aversion questionnaire and got paid.

Stage 1: The Initial Questionnaire

In the first stage, and before receiving the instructions for the successive tasks of the experiment, subjects were asked to fill in the State-Trait Anxiety Inventory (STAI) questionnaire by Spielberger (1976). This questionnaire aims to assess the current anxious state (State-Anxiety) and the proneness to anxiety (Trait-Anxiety) of subjects. It consists of 20 questions on the State-Anxiety and 20 questions on Trait-Anxiety. It assigns two scores, one for State-Anxiety and another for Trait-Anxiety; for each subject there are therefore two measurements from 0 to 80.[8] The score therefore enabled us to have measures for initial anxiety.

[4] For a survey on contest experiments, see Dechenaux et al. (2012).

[5] One specific problem is that, generally, it is hard to disentangle performance from effort in real effort tasks. Performance is observable and is a function of effort (and ability). Effort and ability are unobservable, though limited proxies may be identified.

[6] 'Procedural knowledge' is a repertoire of skills, rules and strategies for using knowledge. Experimenters are usually interested in procedural knowledge and, therefore, they are concerned in writing easy and accessible instructions that would allow all subjects to have the knowledge to understand how their decisions affect their payoffs (Camerer and Hogarth, 1999). Real effort task can be also divided between *real effort task without real outcomes*, such as solving arbitrary mathematical problems, and *real effort task with real outcomes*, such as folding letters or fund-raising. It might be claimed that real effort tasks with real outcomes may incentivize subjects to put more effort than real effort tasks without real outcomes do. In our experiment, all the three real effort tasks we use are real effort tasks without real outcomes.

[7] The experimental instructions are in the online appendix. Subjects were randomly assigned to the seats in the lab where partitions between them assured the anonymity and the avoidance of any communication. Instructions were read aloud and the experimenter checked individually the participants' answers to the control questionnaire.

[8] Scores are adjusted according to the number of not answered questions. If there are 3 or more questions not answered for each set of questions (Trait or State), the score for that set is considered missing (Spielberger, 1972).

Stage 2: The Contest and the Effort Tasks

In the second stage of the session subjects received the instructions for a 2-player contest. They were asked to enter a competition against another subject in the room in order to win a prize of 10 pounds. Pairs were fixed but, as no feedback was received between rounds, dynamic collusive strategies were not possible.

The competition was designed as an all-pay auction, a type of contest in which the subject with the highest bid – in terms of performance - wins the prize with probability 1. Both subjects have a positive valuation of the prize. They both know their own valuation and the valuation of their opponent. Subjects choose their effort $e_i \in [0,1)$ simultaneously, and the cost of effort is linear $C(e_i) = e_i$.

Therefore, in the case of two subjects $i = 1, 2$ subject 1 wins with probability

$$p_1(e_1, e_2) = \begin{cases} 1 & \text{if } e_1 > e_2 \\ \frac{1}{2} & \text{if } e_1 = e_2 \\ 0 & \text{if } e_1 < e_2 \end{cases} \qquad (3.1)$$

The probability with which subject 2 wins is $p_2 = 1 - p_1$. Subject 1's payoff is

$$\pi_1 = \begin{cases} P_w - e_1 & \text{if } e_1 > e_2 \\ 5 - e_1 & \text{if } e_1 = e_2 \\ P_l - e_1 & \text{if } e_1 < e_2 \end{cases} \qquad (3.2)$$

where P_w and P_l are the prizes given to the winner and to the loser of the tournament respectively. In our experiment P_w was equal to £10 and P_l to zero pounds. To simplify, in what follows we talk simply of prize to refer to the £10 prize.

Subjects' effort in this experiment is given by the performance in the effort task. Four different effort tasks were used in four different treatments, the slider task treatment (n = 50), the maths task treatment (n = 54), the grid task treatment (n = 52) and the induced value task treatment (n = 50).[9]

In the *Slider Task* subjects saw forty-eight sliders on their computer screen, with a range of integer values from 0 to 100, initially at 0.[10] They were asked to use their mouse to position the slider at 50. Once the slider was dropped, the value of its position was shown to the right. If not 50, subjects could adjust the position again. There was no limit on the number of attempts per each slider. Subjects had two minutes to correctly

[9] Treatments were run on a between-subject basis. The number of subjects per treatment is slightly different because some students did register for the session but they did not show up. The average payment per subject was 10 pounds and the payment was based on subjects' performance.

[10] Sliders appeared at the very left position. See Appendix 2 for a screen shot (Figure 1).

position as many sliders as they could. Subjects' performance was measured by the number of sliders correctly positioned in two minutes.[11]

For the *Maths Task* subjects were asked to sum a series of four 2-digit numbers on their screen[12]. Examples of possible sums were 21 + 23 + 45 + 67 + 88 and 82 – 13 + 49 + 72 – 15. They were asked to solve as many sums as they could in the time constraint (2 minutes). The number of correct answers measured the subject's performance.

For the *Grid Task* a 5 by 5 grid with randomly distributed 0s and 1s appeared on the screen and subjects were asked to count the number of 1s. They had two minutes to solve as many grids as they could[13]. In this task, which is similar to the one used by Abeler et al. (2011), subjects' performance was measured by the number of grid correctly solved.

In the *Induced-Value Effort Task*, subjects were endowed with ten experimental points and they had to decide how much to invest and how much to keep for themselves[14]. The part of endowment not invested (if any) was converted in pounds at the rate of £1 per point and paid to subjects at the end of the experiment. Subjects with the highest performance, measured here in terms of investment, won the £10 prize (one prize per pair)[15].

In each treatment subjects repeated the effort task for ten rounds. They did not receive any feedback between rounds. At the end of the session one round was randomly selected for the payment. Subjects who achieved the highest performance in each pair won the prize. If subjects in a pair had the same level of performance, the prize was randomly assigned by the computer to one of the two.

Stage 2: Emotions Elicitation

A Visual Analogue Scale (VAS) was used to measure anxiety arousal (Wewers and Lowe, 1990).[16] A bar appeared on the screen whose value 0 could be moved to indicate the mood evaluation "not anxious at all" and value 100 the mood evaluation "very much anxious". Data on arousals of other 7 emotions was also used in order not to focus subjects' minds on anxiety: specifically, tiredness, despondency, frustration, apprehension, tiredness, happiness, anger (the same list as in Slyker and McNally, 1991). The VAS measurement took place before the first real effort task, after 5 rounds, and after all 10.

[11] Guidance and practical advice of the use of this task have been followed.
[12] Participants could see on the screen only one question per time. After submitting their answer, they could proceed to the next question.
[13] This task is similar to the task by Abeler et al. (2011). Their task consisted of a series of 10 by 15 grids filled in randomly by numbers zeros and ones. Subjects were asked to count the number of zeros in each grid. No time limit was imposed.
[14] Invested amounts could be up to two decimal points.
[15] Subjects received a show-up fee of £3 in the sessions where a real effort task was applied (Treatment A, B and C) in order to align the average payment across treatments.
[16] The Visual Analogue Scale is widely used to evaluate subjective phenomena such as mood, anxiety and depression (Scott and Huskisson, 1976; Huskisson, 1983; Downie et al., 1978). It is a quicker and more straightforward tool than the State Anxiety questionnaire, which is instead suitable to measure anxiety arousal just before the beginning of the effort task.

Stage 3: Risk Elicitation and Payment

In the third stage we used a standard Holt and Laury (2002) questionnaire to measure risk aversion. The risk aversion measure can take values from 0 to 9. The higher the number the more risk averse the subjects. At the end of the experiment, a computer screen showed to subjects some information on their payment: the outcome of the lottery, the selected round of the effort task, the outcome of the contest and hence their final earnings. Table 1 shows the overall experimental sequence.

Table 1: Experimental Sequence

STAGE 1	**STAGE 2**	**STAGE 3**
STAI Questionnaire		
Instructions - Practice Round		
	Emotion Elicitation (VAS)	
	5 rounds Effort Task	
	Emotion Elicitation (VAS)	
	5 rounds Effort Task	
	Emotion Elicitation (VAS)	
		Risk-attitude measurement
		Results and Payment

4. Results

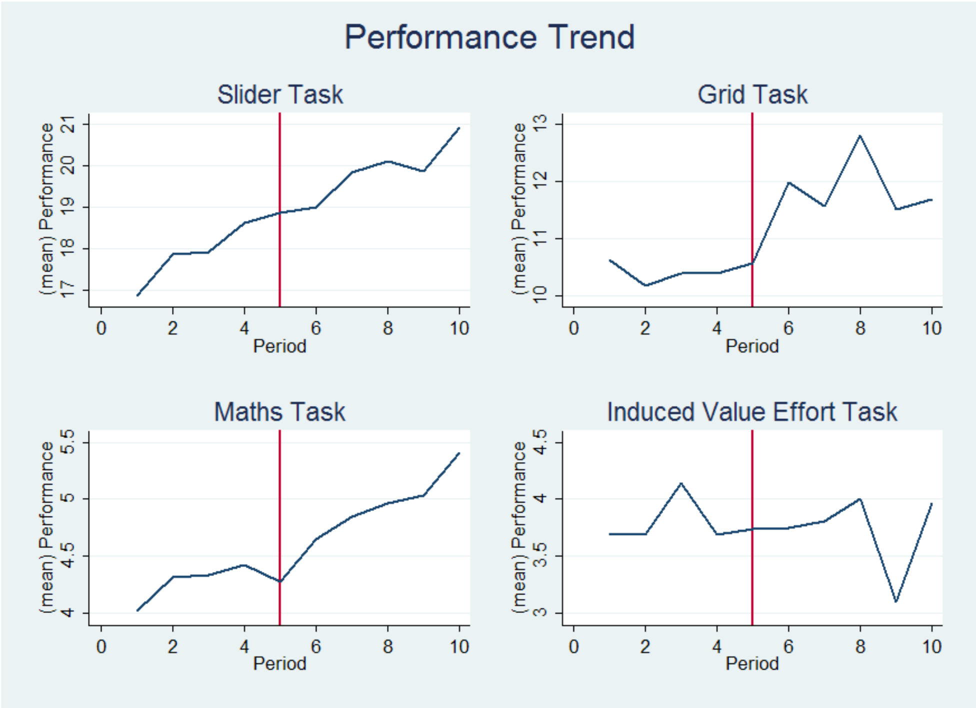
Performance

Table 2 provides some descriptive statistics about performance in the different tasks.[17] Figure 1 and Table 2 show that performance increases over time in real effort task treatments, indicating some learning effects, while this is not the case for the induced value task.[18] In the rest of this section we consider the effect of anxiety, risk attitude and gender on performance.

[17] More descriptives, relative to those presented in this section, are included in the online appendix.

[18] In comparing the last five rounds with the first five, Wilcoxon p = 0.000, 0.001, 0.006 and 0.977 in relation to the slider task, maths task, grid task and induced value task, respectively.

Figure 1: Performance Trend



Notes: The vertical line in the figure indicates the fifth round

Table 2: Aggregate Descriptive Statistics of Performance

	Obs	Mean	Std.Dev	Min	Max
Rounds 1 – 5					
Slider Task	250	18.04	4.329	6	28
Maths Task	270	4.27	1.953	0	9
Grid Task	260	10.43	6.059	0	25
Induced Value Task	250	3.79	3.595	0	10
Rounds 6 – 10					
Slider Task	250	19.95	4.532	9	34
Maths Task	270	4.98	2.303	0	12
Grid Task	260	11.91	6.031	0	25
Induced Value Task	250	3.73	3.719	0	10

Notes: Performance is measured in terms of number of sliders placed at 50 (slider task), correct sums (maths task), number of 1s correctly counted in grids (grid task) and investment placed (induced value task).

In the slider task (Treatment A) the maximum number of sliders that participants can correctly position is 48 per round; they can decide to not do the task at all (so that the minimum is 0). However, at least 6 and at most 32 sliders were correctly positioned in 2 minutes. The maths task and the grid task seem to be more difficult tasks: at least one subject could not solve any mathematical question or grid. There is no upper limit in questions/grids that could be solved for the maths task and for the grid task: the computer could generate an infinite number of grids or mathematical questions in 2 minutes. The best that subjects could do was 12 for the maths ask and 25 for the grid task. Performance (or investment) in the induced value effort task has a much defined range value because subjects receive an endowment of 10 experimental points and they cannot neither invest more than that nor invest a negative amount. The modal range value for the investment decision is between 3.5 and 4. However, some subjects decided to invest everything or not at all (see Table 4).

Anxiety

Table 3 provides details on mean anxiety[19] and risk aversion and on performance by gender. Mean anxiety is lower for subjects that faced the induced value effort task (Treatment D) compared to subjects that carried out the other tasks in both sets of rounds.[20] Moreover, the mean anxiety is significantly higher for subjects that carried out the mathematical task (Treatment B) compared to subjects that faced the other tasks in the second set of rounds.[21] Differently from the other treatments, the level of anxiety in Treatment A decreases significantly in the second set of five rounds.[22]

Table 3: Mean Anxiety, Performance by Gender, and Risk Aversion

	Slider Task Treatment	Maths Task Treatment	Grid Task Treatment	Induced Value Task Treatment
Rounds 1 – 5				
Anxiety	45.12	56.07	49.67	26.42
Performance - Male	19.508	4.6	10.928	3.246
Performance - Female	16.685	4.067	9.963	4.341
Rounds 6 – 10				
Anxiety	40.20	59.41	43.21	24.88
Performance - Male	21.7	5.257	12.168	3.189
Performance - Female	18.323	4.806	11.681	4.264
Risk Aversion	5.38	5.91	5.5	6.12

Measured risk attitude

Table 3 shows that a slightly higher value of risk aversion when this was measured after the maths task treatment and the induced value task treatment. However, only in the induced value treatment is this statistically significant from that in the maths task treatment. Intuitively, the induced value task is about risk

[19] We consider the mean anxiety arousal assessed by the Visual Analogue Scale before the first round and before the 6th round.
[20] Mann-Whitney ranksum test p = 0.004, 0.000, 0.000 for treatment D vs treatment A, B and C respectively for the first set of rounds, and p = 0.014, 0.000, 0.003 for treatment D vs treatment A, B and C respectively for the second set of rounds. All p values in this paper are two-sided.
[21] Mann-Whitney ranksum test p = 0.003, 0.007, 0.000 for treatment B vs treatment A, C and D respectively.
[22] Wilcoxon p = 0.063.

taking in a way that the others are not, and, in a portfolio approach, subjects who have taken risks in the induced value tasks may be less willing to do so in the Holt and Laury task.[23]

Performance by Gender

Men perform significantly better than women in the slider task treatment and in the maths task treatment in both sets of rounds.[24] Mean performance is the same between men and women in the grid task treatment, while women invest more than men in the induced value task treatment in both sets of rounds. There is some initial evidence therefore against Hypothesis 3 and for Hypothesis 4, as it seems not always to be the case that women perform less well in contests.[25]

Regressions on Subjects' Performance

Table 4 shows the estimates of Ordinary Least Squares (OLS) regression models with error clustering on the aggregate level of performance[26] of the first five rounds of the effort task. The independent variables include the VAS anxiety measure collected before the 5 rounds but with full awareness of the task ahead, the Holt and Laury (2002) risk aversion measure, plus controls for initial anxiety (before the task is known) and demographic controls.

Table 4: Regressions with Error Clustering on Performance of the First 5 Rounds

	Slider Task	Maths Task	Grid Task	Ind. Value Task
Anxiety	0.027	-0.017	-0.061**	0.007
Risk Aversion	0.147	-0.135	0.840**	-0.133
Initial State-Anxiety	-0.006	0.028	0.083	0.006
Initial Trait-Anxiety	-0.146**	0.013	-0.096	0.043
Male	2.54**	0.298	-0.112	-0.777
Age	-0.087	0.029	-0.246	0.113
British Nationality	-2.164**	0.422	4.618**	-0.511
School of ECO	-3.457	-0.548	3.552	0.302
Constant	24.029***	3.510	11.855	0.308
Clusters	47	53	52	49
Observations	235	265	260	245
R-Square	0.2381	0.0891	0.2321	0.0773
F – test	3.88	0.92	2.33	1.23

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[23] Wilcoxon p = 0.070 in relation to Treatment D vs. Treatments A. Wilcoxon p = 0.218 in relation to Treatment B vs. Treatments A.
[24] Wilcoxon p = 0.000 for Treatment A and 0.017 for Treatment B for the first set of rounds, and p = 0.000 for Treatment A and 0.054 for Treatment B for the second set of rounds.
[25] Wilcoxon p = 0.157 and 0.328 for Treatment C for the first and second round respectively, and p = 0.019 and 0.022 for Treatment D for the first and second round, respectively.
[26] The use of clusters provides a robust standard error per group (that is, per subject). Sashegyi et al. (2000) argue that where observations over time are taken for different group of subjects an econometric model must control both for intra-cluster correlation and intra-individual correlation within the same cluster.

Notes: Controls include the State-Anxiety and Trait-Anxiety scores, gender (Male = 1 if the subject is male), economics background (School of ECO = 1 if the subject is enrolled in the School of Economics at the University of East Anglia), nationality (British Nationality = 1 if the subject is British) and age.

Controlling for initial anxiety,[27] and against Hypothesis 1, a higher level of task-elicited anxiety negatively affects subjects' performance in the treatment in which the grid task was applied (Treatment C): an increase in anxiety level reduces performance. Also against Hypothesis 1, anxiety does not seem to affect performance with the other effort tasks. Overall, the effect of anxiety on performance is task-specific, as per Hypothesis 4.

Risk aversion is strongly significant in the grid task. It positively affects subjects' performance. In the other treatments, however, this variable does not show a significant coefficient. The evidence therefore provides no support for Hypothesis 2, and clearer support for Hypothesis 4 about task specificity of the effect.

Male subjects perform better than female subjects in the slider task as predicted by Hypothesis 3, but not in the other treatments, which again supports Hypothesis 4.[28]

The same group of independent variables (anxiety, risk aversion, State-Anxiety score and Trait-Anxiety score, and demographic variables) are used to estimate the coefficients of OLS regressions with error clustering on the level of performance in the second five rounds. In these new models we consider the level of elicited anxiety (VAS score) submitted by subjects at the beginning of sixth round. Table 5 shows the coefficients of the regression models. Risk aversion is no longer statistically significant in any task. There is now a negative effect of anxiety on performance in both the maths task and the grid task, but still not in the other two; and the effect of gender is stronger in the slider task, but there remains none elsewhere.

We can summarize the overall results as follows:[29]

RESULT 1. Against Hypothesis 1 but consistent with Hypothesis 4, anxiety negatively affects performance in the grid task, and does not affect performance in the other task.

RESULT 2. Against Hypothesis 2 but consistent with Hypothesis 4, risk aversion positively affects performance (in the first set of rounds) in the grid task, and does not predict performance otherwise.

RESULT 3. As predicted by Hypothesis 3, male subjects achieve higher level of performance than female subjects in slider tasks contests. However, there is no statistically significant effect in the other tasks. Overall, the evidence is supportive of Hypothesis 4.

[27] We find a task-specific negative effect of trait anxiety on performance in the slider task only. This significant effect does not carry over to the second set of 5 rounds analysis (Table 4).

[28] Among our demographic controls, the only partially significant one relates to nationality, but in different directions depending on the task.

[29] Lezzi (2014) presents additional regressions using all seven emotions as independent variables, and finds further support for task specificity of effects on performance.

Table 5: Regression with Error Clustering on Performance of the Second Set of 5 Rounds

	Slider Task	Maths Task	Grid Task	Ind. Value Task
Anxiety	0.000	-0.026**	-0.046**	0.023
Risk Aversion	0.219	-0.092	0.459	-0.171
Initial State-Anxiety	0.005	0.033	-0.011	-0.004
Initial Trait-Anxiety	-0.09	0.015	-0.051	0.044
Male	23.268**	0.202	-0.914	-0.691
Age	0.020	0.144	-0.137	0.134
British Nationality	-2.630**	0.948	4.447**	-0.633
School of ECO	-0.741	-0.828	5.263*	-0.719
Constant	24.031***	1.686	13.794*	0.101
Clusters	47	53	52	49
Observations	235	265	260	245
R-Square	0.2438	0.1264	0.1704	0.1180
F - test	3.87	1.92	1.32	1.66

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5. Discussion and Conclusions

Effort tasks are widely applied in a wide array of economic experiments in an attempt to model natural world effort, such as in labour market experiments and in contests. In our experiments subjects are asked to perform a task where they choose or exert real effort in three different ways. Hence, the task can be an induced value task or one of three different real effort tasks. Our key finding was the task-specificity of our effects. Anxiety negatively affected effort and (at least initially) risk aversion mattered with the grid task, but not otherwise. Gender mattered for the slider task, but not otherwise.

As reviewed in section 2, several factors might be responsible for differences with specific methods. For example, the introduction of a real effort task can trigger heuristics developed in everyday experience which might not be present when subjects are asked to make a choice within a list of numbers.

Surprisingly, no regular check of the robustness of the laboratory results is usually carried out. A single task has typically been implemented in each study and little has been said about the implications of different abilities and skills required in each task. An exception is Brüggen and Strobel (2007): they compared the mathematical task with an induced value task in a gift-exchange game. They examined the differences between the effect of levels of wages on individuals' effort in the setting of a chosen effort (induced effort task) and a real effort (real effort task). They found that individuals reciprocate to higher level of wages exerting more effort and therefore they conclude that the two laboratory methods bring similar results. However, they carefully highlight the presence of greater variance under real effort and that the reciprocal behaviour of individuals is understated under chosen effort. Subjects in fact achieve on average a level of performance four times higher when they solve the mathematical task.

It is well known in the literature on mood inducement techniques that mathematical tasks are anxiety inducing. Psychologists use this task in the laboratory in order to provoke an anxious mood (Keogh and French, 2001; Shostak and Peterson, 1990; Ashcraft and Kirk, 2001). Our paper shows that the grid task might bring subjects to an anxious status as well as the maths task. The mathematical task, very often used in economic experiments, and the grid task do not appear to be equivalent either to the slider task or to the chosen effort task. In the case of the grid task, there is also a decrease in performance.

Experimental work looking at risk aversion and performance has typically been within an induced value paradigm, such as Miller and Pratt (1991). In our experiment we are unable to replicate their finding on the positive effect of risk aversion on induced value effort, but we do find an opposite effect of real effort on performance in the case of the grid task. Our interpretation is that playing safe here is perceived as putting in less effort, though this effect appears real effort task-specific itself.

The task-specificity of the gender effect is also interesting, and naturally fitting with a traditional literature on gender differences or lack thereof due to the nature of the task (e.g. Gneezy et al., 2003).[30] We are only able to replicate the finding of greater effort by men in competitive environments (e.g., Croson and Gneezy, 2009) in the context of slider tasks.

It is important to point out that our discussion on the variables of this experiment does not refer to the magnitude of their impact on subjects' level of performance. We are not interested in either measuring the relative effects of each variable through treatments or proposing the advantage of one type of task over another. Our claim is simply that the significance of the independent variables considered on performance strongly depends on the effort task used and that experimental results may be driven by the effort elicitation method being employed.

[30] In our experiment we can rule out the possibility that performance was affected by any stereotype threat linked to the type of the effort task. We did not make salient any group identity and neither the experimental instructions nor the decision environment faced by subjects refer to gender.

References

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

Ashcraft, M. H. and E. P. KIrk (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology 130* (2), 224.

Azar, O. H. (2009). Does relative thinking exit in mixed compensation schemes? Technical report, *Working paper*, Ben-Gurion University of the Negev.

Balafoutas, L. and M. Sutter (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science 335* (6068), 579–582.

Bosman, R. and F. van Winden (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal 112* (476), 147–169.

Bosman, R., M. Sutter, and F. van Winden (2005). The impact of real effort and emotions in the power-to-take game. *Journal of Economic Psychology 26* (3), 407–429.

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

Breitmoser, Y., J. Tan, and D.J. Zizzo (2010). Understanding perpetual R&D races. *Economic Theory 44* (3), 445–467.

Brüggen, A. and M. Strobel (2007). Real effort versus chosen effort in experiments. *Economics Letters 96* (2), 232–236.

Bull, C., A. Schotter, and K. Weigelt (1987). Tournaments and piece rates: an experimental study. *Journal of Political Economy 95* (1), 1–33.

Burrows, P. and G. Loomes (1994). The Impact of fairness on bargaining. *Empirical Economics 19* (2), 201–221.

Camerer, C. F. and R. M. Hogarth (1999). The effects of financial incentives in experiments: A review and capital-labour-production framework. *Journal of Risk and Uncertainty 19* (1), 7–42.

Carpenter, J., Matthews, P. H., & Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *The American Economic Review*, *100*(1), 504-517.

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

Charness, G. and M. Villeval (2009). Cooperation and competition in intergenerational experiments in the field and the laboratory. *The American Economic Review 99* (3), 956–978.

Charness, G., G. R. Frechette, and J. H. Kagel (2004). How robust is laboratory gift exchange? *Experimental Economics 7* (2), 189–205.

Croson, R. and U. Gneezy (2009). Gender diferences in preferences. *Journal of Economic Literature*, 448–474.

Dechenaux, E., Kovenock, D., & Sheremeta, R. M. (2012). A survey of experimental research on contests, all-pay auctions and tournaments (No. SP II 2012-109). *Discussion Paper, Social Science Research Center Berlin (WZB)*, Research Area'Markets and Politics', Research Professorship & Project'The Future of Fiscal Federalism'.

Dickinson, D. (2001). The carrot vs. the stick in work team motivation. *Experimental Economics 4* (1), 107–124.

Dickinson, D., & Villeval, M. C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic behavior*, *63*(1), 56-76

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

Downie, W. W., Leatham, P. A., Rhind, V. M., Pickup, M. E., & Wright, V. (1978). The visual analogue scale in the assessment of grip strength. *Annals of the rheumatic diseases*, *37*(4), 382-384.

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, *16*(6), 679-688.

Eysenck, M. and M. Calvo (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion 6*, 409–434.

Fahr, R. and B. Irlenbusch (2010). Fairness as a constraint on trust and reciprocity. *Economics Letters 66*, 275–282.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labour Economics 24* (1), 39–57.

Falk, A., Huffman, D., & Sunde, U. (2006). *Self-confidence and search* (No. 2525). IZA Discussion Papers.

Fehr, E. and K. M. Schmidt (2004). Fairness and Incentives in a multi-task principal–agent model. *The Scandinavian Journal of Economics 106* (3), 453–474.

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

Fehr, E., G. Kirchsteiger, and A. Riedl (1993). Does fairness prevent market clearing? An experimental Investigation. *The Quarterly Journal of Economics 108* (2), 437–459.

Fehr, E., S. Gächter, and G. Kirchsteiger (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, 833–860.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economics experiments. *Experimental Economics 10* (2), 171–178.

Gächter, S. and A. Falk (2002). Reputation and reciprocity: Consequences for the labour relation. *The Scandinavian Journal of Economics 104* (1), 1–26.

Gächter, S. and C. Thöni (2010). Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behaviour & Organization 76* (3), 531–543.

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

Gneezy, U. (2002). Does high wage lead to high profits: An experimental study of reciprocity using real efforts. *Graduate School of Business, University of Chicago*.

Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *The American Economic Review 94* (2), 377–381.

Gneezy, U. and J. List (2006). Putting behavioural economics to work: Testing for gift exchange in labour markets using field experiments. *Econometrica 74* (5), 1365–1384.

Gneezy, U., M. NIederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118* (3), 1049–1074.

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

Harbring, C. and B. Irlenbusch (2005). Incentives in Tournaments with Endogenous Prize Selection. *Journal of Institutional and Theoretical Economics JITE 161* (4), 636–663.

Harbring, C. and G. K. Lünser (2008). On the competition of asymmetric agents. *German Economi Review 9* (3), 373–395.

Hennig-Schmidt, H., Sadrieh, A., & Rockenbach, B. (2010). In search of workers' real effort reciprocity—a field and a laboratory experiment. *Journal of the European Economic Association*, *8*(4), 817-837.

Heyman, J. and D. Ariely (2004). Effort for payment: A tale of two markets. *PsychologIcal Science 15* (11), 787–793.

Hillman, A. L. and E. Katz (1984). Risk-averse rent seekers and the social cost of monopoly power. *The Economic Journal 94* (373), 104–110.

Hoffman, E. and M. Spitzer (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *Journal of Legal Studies 14*, 259.

Hoffman, E., K. McCabe, K. Shachat, and V. Smith (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behaviour 7* (3), 346–380.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *The American economic review 92* (5), 1644–1655.

Huskisson, E. C. (1983). *Visual analogue scales*. Pain measurement and assessment. (Vol. 33, No. 7). New York: Raven Press.

Ivanova-Stenzel, R., & Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, *32*(5), 797-808.

Keogh, E. and C. C. French (2001). Test anxiety, evaluative stress, and susceptibility to distraction from threat. *European Journal of Personality 15* (2), 123–141.

Konow, J. (2000). Fair shares: Accountability and cognitive dissonance in allocation decisions. *The American Economic Review 90* (4), 1072–1091.

Konow, J., Saijo, T., & Akai, K. (2009). Morals and mores: Experimental evidence on equity and equality. *Loyola Marymount University*.

Konrad, K. A. and H. Schlesinger (1997). Risk aversion in rent-seeking and rent-augmenting games. *The Economic Journal 107* (445), pp. 1671–1683.

Kube, S., M. Maréchal, and C. Puppe (2006). Putting reciprocity to work: Positive versus negative responses in the field. *University of St. Gallen Economics Discussion Paper* (2006-27).

Lévy-Garboua, L., D. Masclet, and C. Montmarquette (2009). A behavioural Laffer curve: Emergence of a social norm of fairness in a real effort experiment. *Journal of Economic Psychology 30* (2), 147–161.

Lezzi, E. (2014). Agents' Performance and Emotions: An Experimental Investigation. PhD Thesis. University of East Anglia, UK.

Millner, E. L. and M. D. Pratt (1989). An experimental investigation of effIcient rent-seeking. *Public Choice 62* (2), 139 – 151.

Millner, E. L. and M. D. Pratt (1991). Risk aversion and rent-seeking: An extension and some experimental evidence. *Public Choice 69* (1), 81–92.

Montmarquette, C., J. Rullière, M. Villeval, and R. Zeiliger (2004). Redesigning teams and Incentives in a merger: An experiment with managers and students. *Management Science 50* (10), 1379–1389.

Morgan, J., H. Orzen, and M. Sefton (2012). Endogenous entry in contests. *Economic Theory 51* (2), 435–463.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review Economics. 3* (1), 601–630.

Ong, D. and Z. C. Chen (2012). Tiger women: An all-pay auction experiment on gender signalling of desire to win. *Available at SSRN 1976782* .

Pokorny, K. (2008). Pay but do not pay too much: An experimental study on the impact of incentives. *Journal of Economic Behaviour & Organization 66* (2), 251–264.

Price, C. and R. Sheremeta (2012). Endowment origin, demographic effects and individual preferences in contests. *Demographic Effects and Individual Preferences In Contests (March 28, 2012)*.

Sashegyi, A. I., K. S. Brown, and P. J. Farrel (2000). Application of a generalized random effects regression model for cluster-correlated longitudinal data to a school-based smoking prevention trial. *American Journal of Epistemology 152* (12), 1192–1200.

Schotter, A. and K. Weigelt (1992). Asymmetric tournaments, equal opportunities laws, and affirmative action: Some experimental results. *The Quarterly Journal of Economics 107* (2), 511–539.

Scott, J. and Huskisson, E. C. (1976). Graphic representation of pain. *Pain*, *2*(2), 175-184.

Sheremeta, R. M., & Wu, S. Y. (2012). *Testing canonical tournament theory: On the impact of risk, social preferences and utility structure* (No. 6304). Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit.

Shostak, B. B. and R. A. Peterson (1990). Effects of anxiety sensitivity on emotional response to a stress task. *Behaviour Research and Therapy 28* (6), 513–521.

Skaperdas, S. and L. Gan (1995). Risk aversion in contests. *The Economic Journal*, 951–962.

Sloof, R. and C. van Praag (2008). Performance measurement, expectancy and agency theory: An experimental study. *Journal of Economic Behaviour & Organization 67* (3), 794–809.

Slyker, J. P. and R. McNally (1991). Experimental Induction of anxious and depressed mood: Are Velten and musical procedures necessary? *Cognitive Therapy and Research 15* (1), 33–45.

Smith, V. (1976). Experimental economics: Induced value theory. *American Economic Review 66* (2), 274–279.

Smith, V. (1979). Incentive compatible experimental processes for the provision of public goods. *Research in experimental economics*.

Spielberger, C. D. (1972). *State-Trait Anxiety Inventory*. John Wiley and Sons, Inc.

Spielberger, C. D. (1976). The nature and measurement of anxiety. In C. D.Spielberger and R. Diaz-Guerrero (Eds.), *Cross-cultural Anxiety*. DC: Hemisphere: Washington.

Sutter, M. and H. Ick-Hannemann (2003). Taxation and the veil of Ignorance–a real effort experiment on the Laffer curve. *Public Choice 115* (1), 217–240.

Treich, N. (2010). Risk-aversion and prudence in rent-seeking games. *Public Choice 145* (3-4), 339–349.

van Dijk, F., J. Sonnemans, and F. van Winden (2001). Incentive systems in a real effort experiment. *European Economic Review 45* (2), 187–214.

Vesterlund, L. and M. Nierderle (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics 122* (3), 1067–1101.

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in nursing & health*, *13*(4), 227-236.

Appendix 1. Experimental Instructions

Experimental Instructions - Treatment A - The Slider Task

[The experimenter reads these instructions at the beginning of the experiment].

You are participating in an experiment on decision making. The experiment is expected to last 1 hour.

During the experiment you are not allowed to communicate with other participants. If you have any questions at any moment please raise your hand and the experimenter will come to your desk.

The experiment consists of **THREE** SECTIONS:

FIRST SECTION

In the first section you are asked to fill in a questionnaire. The questionnaire consists of 2 parts: PART A and PART B. For each part you will receive specific instructions. Please read the instructions carefully. After that all of you havde completed the questionnaire, you will receive the instructions for the second and third section of the experiment.

SECOND SECTION

[The experimenter reads these instructions after participants have completed the (STAI) questionnaire].

In the second section you are randomly matched with another participant in this room. You will not be told whom you are matched with. We will call him or her your “co-participant”.

In this section you and your co-participant are asked to perform a task that will be described shortly. In this task you get a score based on your performance.

The task

You will see 48 sliders on the screen. Each slider has integer values from 0 to 100.

Each slider will appear on position 0.

Your task is to position each slider at exactly 50 with the help of your mouse. The sliders can be adjusted and readjusted an unlimited number of times and the current position is displayed to the right of each slider.

Your score

You get a score of one point per each slider you have positioned at 50.

Example:

If you position five sliders at exactly 50, you get a score of 5 points.

You will have two minutes to correctly position as many sliders as you can. We will call these two minutes "round".

In this experiment you are asked to carry out this task for 10 rounds. You will be matched with the same co-participant among the 10 rounds. All the sliders will appear at position 0 at the beginning of each round. Therefore, your score will be reset. This means that scores of previous rounds will not be added to scores of successive rounds.

Your earnings in the second section

At the end of the experiment the computer will randomly select one round out of 10. The selected round will be the same for you and your co-participant.

The computer will compare your score to the score of your co-participant obtained in the selected round.

Three situations can occur:

• If your score in the selected round is higher than your co-participant's score, you earn a prize of £10;

• If your score in the selected round is lower than your co-participant's score, you earn £0;

• If your score in the selected round is equal to your co-participant's score, the computer will randomly select one of you; you have therefore a 50% chance of being selected and of winning the prize.

Example 1

Suppose the computer selects round 1. If your score in round 1 is higher than your co-participant's score, your earnings are £10.00.

Example 2

Suppose the computer selects round 5. If your score in round 5 is lower than your co-participant's score, your earnings are £0.00.

Example 3

Suppose the computer selects round 10. If your score in round 10 is equal to your co-participant's score, your earnings are £10 or £0 with a 50% chance.

THIRD SECTION

In the next screen you will be asked to make some decisions. Your decisions will affect your final earnings: what you will earn from this task will be added to your earnings of the previous task. You now need to make 9 decisions for each of two successive computer screens. Each decision is a paired choice between two options (for example, “Option A” and “Option B”). You will make 9 decisions and record these in the final column, but only one of them from each computer screen will be used in the end to determine your earnings. You will only know which one at the end of this session.

Experimental Instructions - Treatment B - The Maths Task

[The experimenter reads these instructions at the beginning of the experiment].
You are participating in an experiment on decision making. The experiment is expected to last 1 hour. During the experiment you are not allowed to communicate with other participants. If you have any questions at any moment please raise your hand and the experimenter will come to your desk.

The experiment consists of **THREE** SECTIONS:

FIRST SECTION
In the first section you are asked to fill in a questionnaire. The questionnaire consists of 2 parts: PART A and PART B. For each part you wIll receive specific instructions. Please read the instructions carefully.
After that all of you have completed the questionnaire, you will receive the instructions for the second and third section of the experiment.

SECOND SECTION
[The experimenter reads these instructions after participants have completed the (STAI) questionnaire].

In the second section you are randomly matched with another participant in this room. You wIll not be told whom you are matched with. I will call him or her your “co-participant”.
In this section you and your co-participant are asked to perform a task that will be described shortly. In this task you get a score based on your performance.

The task
The task consists of adding-up four numbers that are randomly generated and that are displayed on your screen.
For example, you will see on your screen: 50 + 20 + 10 + 33
Your task is to sum these numbers that is 50 + 20 + 10 + 33.
The correct answer in this example is 113.

The use of paper, pencil or calculator is forbidden. You must make these calculations in your head.
Once your calculation has been done, you enter your answer in the specified area and you submit your answer by clicking the “submit” button. Having submitted your answer a new series of numbers appears automatically on your screen. Therefore, you are asked to sum these new numbers and to submit your answer.

You will have two minutes to submit as many correct answers as you can. I will call these two minutes “round”. In this experiment you are going to perform the arithmetic task for 10 rounds. You will be matched with the same co-participant among the 10 rounds but the series of numbers will be different in each round.

Your score
You get a score of one point per each correct answer you submit. You will not be told whether your answers are correct or wrong when you submit them.

Example:
In the example: 50 + 20 + 10 + 33 the correct answer is 113. If your answer is 113 you get a score of 1 point.

At the beginning of each round your score is reset. This means that scores of previous rounds will not be added to scores of successive rounds. Your score is independent in each round.

Your earnings in the second section
At the end of the experiment the computer will randomly select one round out of 10. The selected round will be the same for you and your co-participant.
The computer will compare your score to the score of your co-participant obtained in the selected round.

Three situations can occur:
• If your score in the selected round is higher than your co-participant’s score, you earn a prize of £10;
• If your score in the selected round is lower than your co-participant’s score, you earn £0;
• If your score in the selected round is equal to your co-participant’s score, the computer will randomly select one of you; you have therefore a 50% chance of being selected and of wining the prize.

Example 1
Suppose the computer selects round 1. If your score in round 1 is higher than your co-participant’s score, your earnings are £10.00.

Example 2
Suppose the computer selects round 5. If your score in round 5 is lower than your co-participant’s score, your earnings are £0.00.

Example 3
Suppose the computer selects round 10. If your score in round 10 is equal to your co-participant’s score, your earnings are £10 or £0 with a 50% chance.

THIRD SECTION
In the next screen you will be asked to make some decisions. Your decisions will affect your final earnings: what you will earn from this task will be added to your earnings of the previous task.
You now need to make 9 decisions for each of two successive computer screens. Each decision is a paired choice between two options (for example, “Option A” and “Option B”). You will make 9 decisions and record these in the final column, but only one of them from each computer screen will be used in the end to determine your earnings. You will only know which one at the end of this session.

Experimental Instructions - Treatment C - The Grid Task

[The experimenter reads these instructions at the beginning of the experiment].
You are participating in an experiment on decision making. The experiment is expected to last 1 hour.
During the experiment you are not allowed to communicate with other participants. If you have any questions at any moment please raise your hand and the experimenter will come to your desk.

The experiment consists of **THREE** SECTIONS:

FIRST SECTION
In the first section you are asked to fill in a questionnaire. The questionnaire consists of 2 parts: PART A and PART B. For each part you will receive specific instructions. Please read the instructions carefully.
After that all of you have completed the questionnaire, you will receive the instructions for the second and third section of the experiment.

SECOND SECTION
[The experimenter reads these instructions after participants have completed the (STAI) questionnaire].

In the second section you are randomly matched with another participant in this room. You will not be told whom you are matched with. I will call him or her your “co-participant”.

In this section you and your co-participant are asked to perform a task that will be described shortly. In this task you get a score based on your performance.

The task
For task 1 you will see one table on the screen. You will be asked to count the number of ones in each table. Then please write down your answer in the box at the bottom of the screen and submit it by clicking the "Submit" button. After you have submitted your answer a new table will be generated and displayed on the screen. Please notice that once you have submitted you cannot change your answer.

ThIs task will last for 2 minutes. Try to solve as many tables as you can within the 2 minutes.
The remaining time is displayed in seconds in the top right corner of the screen.

Your score
You get a score of one point per each table you have correctly solved. *Example:* if you correctly solve five tables, you get a score of 5 points.

You will have two minutes to correctly solve as many tables as you can. I will call these two minutes "round". In this experiment you are asked to carry out this task for 10 rounds. You will be matched with the same co-participant among the 10 rounds. At the beginning of each round your score will be reset. This means that scores of previous rounds will not be added to scores of successive rounds.

Your earnings in the second section
At the end of the experiment the computer will randomly select one round out of 10. The selected round will be the same for you and your co-participant.

The computer will compare your score to the score of your co-participant obtained in the selected round.

Three situations can occur:
• If your score in the selected round is higher than your co-participant's score, you earn a prize of £10;
• If your score in the selected round is lower than your co-participant's score, you earn £0;
• If your score in the selected round is equal to your co-participant's score, the computer will randomly select one of you; you have therefore a 50% chance of being selected and of winning the prize.

Example 1
Suppose the computer selects round 1. If your score in round 1 is higher than your co-participant's score, your earnings are £10.00.

Example 2
Suppose the computer selects round 5. If your score in round 5 is lower than your co-participant's score, your earnings are £0.00.

Example 3
Suppose the computer selects round 10. If your score in round 10 is equal to your co-participant's score, your earnings are £10 or £0 with a 50% chance.

THIRD SECTION
In the next screen you will be asked to make some decisions. Your decisions will affect your final earnings: what you will earn from this task will be added to your earnings of the previous task.
You now need to make 9 decisions for each of two successive computer screens. Each decision is a paired choice between two options (for example, "Option A" and "Option B"). You will make 9 decisions and record these in the final column, but only one of them from each computer screen will be used in the end to determine your earnings. You will only know which one at the end of this session.

Experimental Instructions - Treatment D - The IV Effort Task

[The experimenter reads these instructions at the beginning of the experiment].
You are participating in an experiment on decision making. The experiment is expected to last 1 hour.
During the experiment you are not allowed to communicate with other participants. If you have any questions at any moment please raise your hand and the experimenter will come to your desk.

The experiment consists of **THREE** SECTIONS:

FIRST SECTION
In the first section you are asked to fill in a questionnaire. The questionnaire consists of 2 parts: PART A and PART B. For each part you will receive specific instructions. Please read the instructions carefully.
After that all of you have completed the questionnaire, you will receive the instructions for the second and third section of the experiment.

SECOND SECTION
[The experimenter reads these instructions after participants have completed the (STAI) questionnaire].

In the second section you are randomly matched with another participant in this room. You will not be told whom you are matched with. I will call him or her your "co-participant".

In this section you and your co-participant are asked to perform a task that will be described shortly. In this task you get a score based on your performance.

The task
You are given an endowment of 10 Experimental Points.
You are asked to decide how much you want to invest and how much you want to keep for yourself.

For example, if you decide to invest 5.55 experimental points your remaining endowment will be 10.00 - 5.55 = 4.45 experimental points.

Once you have decided, you enter your answer in the specified area and you submit your decision by clicking the "submit" button.

You will be asked to make this decision 10 times. I will call these times "rounds". Therefore this experiment will have 10 rounds in which you are matched with the same co-participant. You receive an endowment of ten experimental points at the beginning of each round. You will not be told about your co-participant's decisions.

Your earnings in the second section
At the end of the experiment the computer will randomly select one round out of 10. The selected round will be the same for you and your co-participant.

The computer will compare your decision to the decision of your co-participant made in the selected round.

Three situations can occur:

- If your investment in the selected round is higher than your co-participant's investment, you earn a prIze of £10;
- If your investment in the selected round is lower than your co-participant's investment, you earn £0;
- If your investment in the selected round is equal to your co-participant's investment, the computer will randomly select one of you; you have therefore a 50% chance of being selected and of winning the prize.

Your remaining endowment in the selected round will be converted in pounds with the following conversion rate:

0.01 Experimental Points = £0.01

For example, 1 Experimental Point will be converted in £1.

Therefore, your earnings in the second section will be:

- £10 + your remaining endowment in the selected round if your investment in the selected round is higher than your co-participant's investment;
- £0 + your remaining endowment in the selected round if your investment in the selected round is lower than your co-participant's investment;
- £10 + your remaining endowment in the selected round if your investment in the selected round is equal to your co-participant's investment and the computer selects you as the winner;
- £0 + your remaining endowment in the selected round if your investment in the selected round is equal to your co-participant's investment and the computer selects your co-participant as the winner.

Example 1

Suppose the computer selects round 1. Suppose that your investment in round 1is 4.49 experimental points. Your remaining endowment is 10.00 - 4.49 = 5.51 experimental points. Your remaining endowment converted in pounds Is £5.51. If in round 1 your investment is higher than your co-participant's investment, your earnings are £10.00 + £5.51 = £15.51.

Example 2

Suppose the computer selects round 4. Suppose that your investment in round 4 is 3.40 experimental points. Your remaining endowment is 10.00 - 3.40 = 6.60 experimental points. Your remaining endowment converted in pounds is £6.60. If in round 4 your investment is lower than your co-participant's investment, your earnings are £0.00 + £6.60 = £6.60.

Example 3

Suppose the computer selects round 10. 166

Suppose that your investment in round 10 is 5.00 experimental points. Your remaining endowment is 10.00 - 5.00 = 5.00 experimental points. Your remaining endowment converted in pounds is £5.00. If in round 10 your investment is equal to your co-participant's investment, your earnings are £10.00 + £5.00 = £15.00 if the computer selects you as the winner OR £0.00 + £5.00 = £5.00 if the computer selects your co-participant as the winner.

THIRD SECTION

In the next screen you will be asked to make some decisions. Your decisions will affect your final earnings: what you will earn from this task will be added to your earnings of the previous task.

You now need to make 9 decisions for each of two successive computer screens. Each decision is a paired choice between two options (for example, "Option A" and "Option B"). You will make 9 decisions and record these in the final column, but only one of them from each computer screen will be used in the end to determine your earnings. You will only know which one at the end of this session.

Appendix 2. – Tables and Figures

Figure A1: Slider Task screen shot

Figure A2: Grid Task screen shot

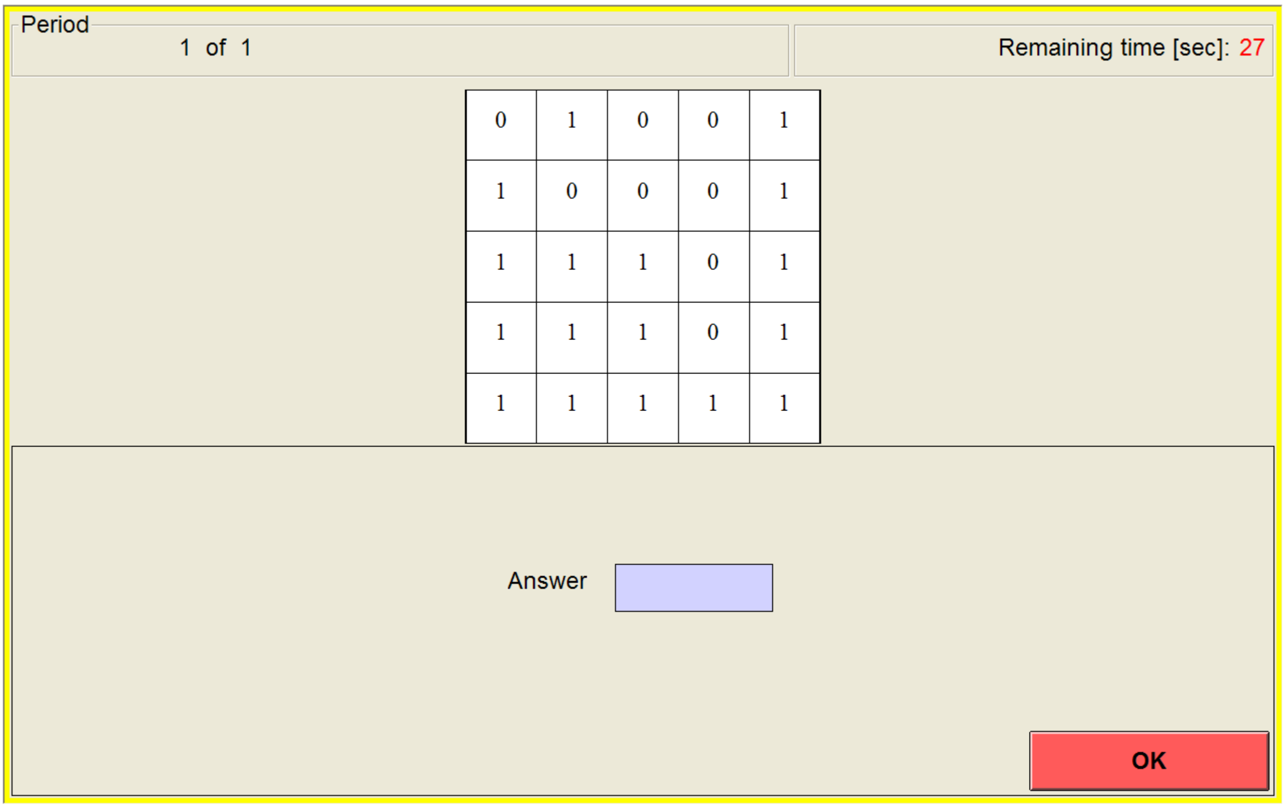


Figure A3: Performance Distributions

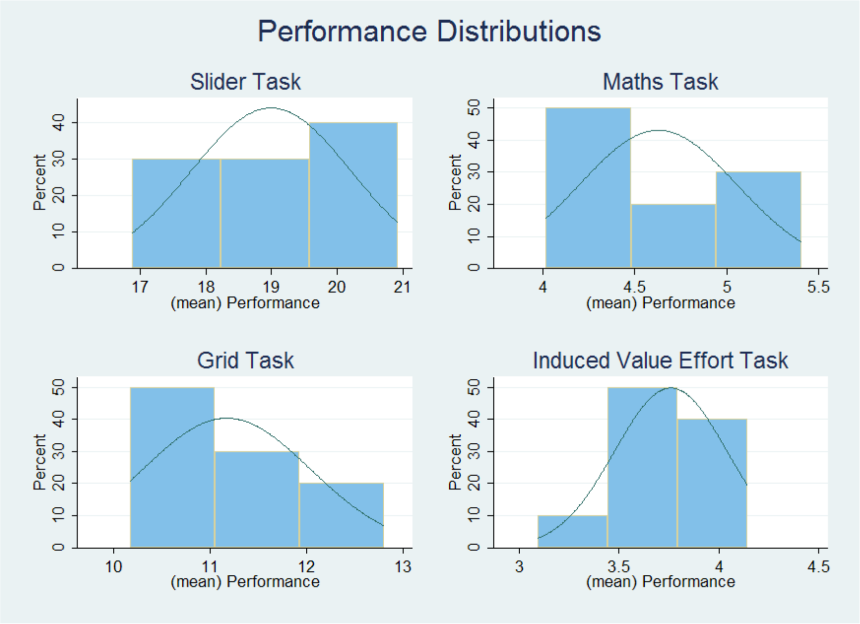


Figure A4: Performance Box Plot

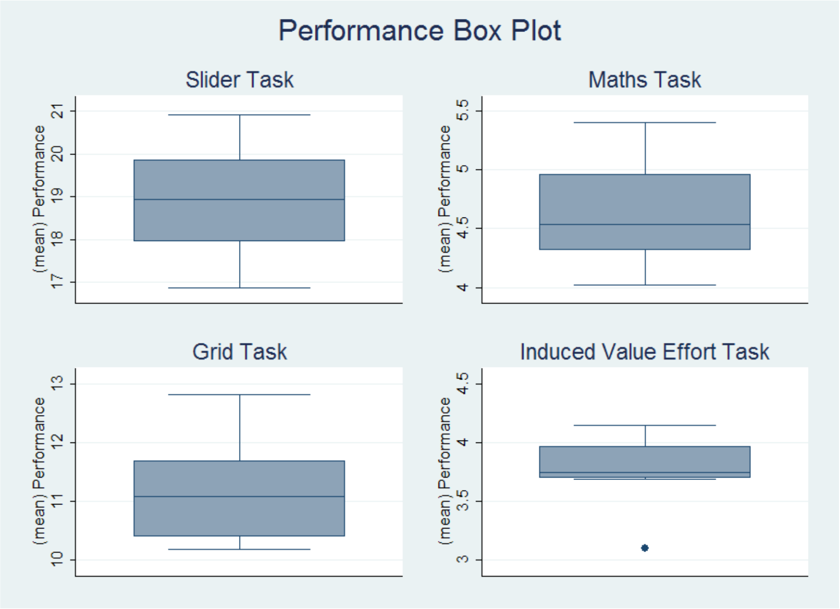


Table A1: Aggregate Descriptive Statistics of Anxiety – by Sets of Rounds

	Obs	Mean	Std. Dev	Min	Max
Rounds 1 - 5					
Slider Task	50	45.12	31.885	0	100
Maths Task	54	56.07	27.53	0	100
Grid Task	52	49.67	30.013	0	100
Ind. Value Effort Task	50	26.42	26.520	0	100
Rounds 6 - 10					
Slider Task	50	40.20	31.784	0	100
Maths Task	54	59.41	31.745	0	100
Grid Task	52	43.21	31.917	0	100
Ind. Value Effort Task	50	24.88	30.569	0	100

Table A2: Summary Statistics of Risk Aversion

	Obs	Mean	Std. Dev	Min	Max
Slider Task	50	5.38	1.794	1	9
Maths Task	54	5.91	2.226	0	9
Grid Task	52	5.5	2.227	0	9
Ind. Value Effort Task	50	6.12	2.067	1	9

Table A3: Summary Statistics of Performance by Gender

		Obs	Mean	Std. Dev	Min	Max
Rounds 1 - 5						
Slider Task	**Male**	120	19.508	3.902	10	28
	Female	130	16.685	4.274	6	28
Maths Task	**Male**	105	4.6	2.003	0	9
	Female	165	4.067	1.897	0	9
Grid Task	**Male**	125	10.928	5.966	0	23
	Female	135	9.963	6.128	0	25
Ind. Value Effort Task	**Male**	125	3.246	3.771	0	10
	Female	125	4.341	3.336	0	10
Rounds 6 - 10						
Slider Task	**Male**	120	21.7	4.351	11	34
	Female	130	18.323	4.083	9	27
Maths Task	**Male**	105	5.257	2.171	1	10
	Female	165	4.806	2.373	0	12
Grid Task	**Male**	125	12.168	6.149	0	24
	Female	135	11.681	5.933	0	25
Ind. Value Effort Task	**Male**	125	3.189	3.706	0	10
	Female	125	4.264	3.669	0	10