I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17287

One Size Fits All? The Interplay of Incentives, Effort Provision, and Personality

Zvonimir Bašić
Stefania Bortolotti
Daniel Salicath
Stefan Schmidt
Sebastian O. Schneider
Matthias Sutter

SEPTEMBER 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17287

One Size Fits All? The Interplay of Incentives, Effort Provision, and Personality

Zvonimir Bašić
University of Glasgow

Stefania Bortolotti
University of Bologna and IZA

Daniel Salicath
Norwegian Labour and Welfare Administration

Stefan Schmidt
Max Planck Institute for Research on Collective Goods

Sebastian O. Schneider
Max Planck Institute for Research on Collective Goods

Matthias Sutter
Max Planck Institute for Research on Collective Goods, University of Cologne, University of Innsbruck, University of Venice Ca' Foscari and IZA

SEPTEMBER 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

One Size Fits All? The Interplay of Incentives, Effort Provision, and Personality*

Incentives are supposed to increase effort, yet individuals react differently to incentives. We examine this heterogeneity by investigating how personal characteristics, preferences, and socio-economic background relate to incentives and performance in a real effort task. We analyze the performance of 1,933 high-school students under a Fixed, Variable, or Tournament payment. Productivity and beliefs about relative performance, but hardly any personal characteristics, play a decisive role for performance when payment schemes are exogenously imposed. Only when given the choice to select the payment scheme, personality traits, economic preferences and socioeconomic background matter. Algorithmic assignment of payment schemes could improve performance, earnings, and utility, as we show.

JEL Classification: C93, D91, J24, J41

Keywords: effort, productivity, incentives, personality traits, preferences, socio-economic background, ability, heterogeneity, sorting, algorithm, lab-in-the-field experiment

Corresponding author:
Matthias Sutter
Max Planck Institute for Research on Collective Goods
Kurt-Schumacher-Str. 10
53113 Bonn
Germany

E-mail: matthias.sutter@coll.mpg.de

* This study has been pre-registered at the AEA RCT Registry under AEARCTR-0008360 and has obtained ethical approval (IRB of the University of Innsbruck, #07/2019). For helpful discussions and comments we thank Alexander Cappelen, Uri Gneezy and Bertil Tungodden as well as seminar participants at the 8th M-BEPS and the meeting of the European Economic Association in Barcelona.

1 Introduction

Understanding how to motivate people to provide effort is of key importance for success in many domains of life, ranging from the educational sector to the labor market. Extensive theoretical and empirical research has focused on the role of both intrinsic and extrinsic incentives to promote effort provision (Prendergast, 1999; Gneezy et al., 2011; Gneezy and Rey-Biel, 2014; Gneezy et al., 2019; Cassar and Meier, 2018). Yet, it is still relatively poorly understood which incentives are best for which people, as humans react in very different ways to the same incentives. Therefore, it has become a major issue in management to understand the heterogeneity of effort provision in reaction to different incentives (Opitz et al., 2024). Some people thrive and express their best potential in competitive environments, while some instead choke under such pressure (Dohmen, 2008). Some people are diligent and work hard regardless of the environment, while others need monetary rewards to be motivated. Scientific evidence offers surprisingly little guidance in understanding the underpinning of this heterogeneity.

In this paper, we examine how a broad set of personal characteristics, skills and preferences, as well as one's socio-economic background shape performance under various payment schemes. A better understanding of what motivates individuals to put in strong effort under different payment schemes would have far-reaching implications for practitioners as well as for theoretical models. However, field data on personal characteristics and productivity under different incentive contracts are scarce, either because companies do not have data on personality traits or, if they do, it is proprietary to combine these data with information about payment schemes and productivity. Even in the latter case, a proper identification strategy is difficult to achieve because of self-selection as well as market frictions. It is likely that individuals choose different career paths conditional on the incentives offered in the different paths. On top of that, once working under specific incentives, this experience may affect a person's reaction to different payment schemes. For instance, a competitive environment might shape how one reacts to a tournament payment. In order to mitigate all of these issues, we conducted a controlled large-scale lab-in-the-field experiment in a setting with very limited self-selection and where participants have hardly any workplace experience with different payment schemes.

We ran our experiment in German high schools, with students from grade ten and older (sixteen to twenty years of age, average of 17.1 years). Relying on a pool of high school students offers a number of advantages, which are crucial for our research question. First, they are highly heterogeneous in terms of socio-economic background, traits, and ability levels, since we targeted schools called 'Gesamtschule" in Germany, which are schools that commonly comprise both low and high education tracks within the same institution, thus encompassing students from all segments of society. This is an advantage of our sample.

Other samples – like university students or workers from a particular industry – would typically be much less heterogeneous than a high school sample from Germany where a large majority of a birth cohort attends high school.[1] Second, high-school students have usually not yet experienced different compensation schemes through experiences in the labor market; rather, they are all exposed to similar incentives in schools. As the mentioned experience may shape preferences, this raises a concern about recruiting individuals who already sorted into jobs. Such potential concerns are void in our study. Third, students are close to entering the labor market – either as full-time workers after high school or as part-time workers during tertiary education after high school. This means that our sample, while not yet exposed to (substantial) labor market experience, will soon work under different incentives in labor markets.

In our experiment, we implemented a tedious counting task adapted from Abeler et al. (2011) to measure effort provision. We used three different incentive schemes, called "Fixed", "Variable", and "Tournament" payment. The Fixed payment pays a flat wage for doing the real effort task. The Variable payment offers a piece rate per correctly solved task, and the Tournament payment offers a higher piece rate than in "Variable" if a subject performed better than another person, but a lower piece rate otherwise. We chose three payment schemes that cover a large share of contracts actually offered on labor markets. We implemented two treatments. In one treatment, we exogenously assigned participants to one of the three different payments schemes. In the other treatment, we let participants themselves choose which payment scheme they preferred. The goal of the latter treatment is threefold. First, it allows us to test if participants are able to self-select into the payment scheme that maximizes their potential monetary earnings. Second, it serves as a test to see if having agency over the type of payment itself has an effect on one's performance. Third, we can examine whether the same personal characteristics are relevant for sorting into different payment schemes and for being productive with a given payment scheme, or whether the two aspects – choice of payment scheme, and performance in given payment scheme – are driven by different traits and characteristics.

With regards to personal characteristics and traits, we consider a plethora of factors that have been proven important in understanding labor market outcomes, such as socio-economic status (see, e.g., Heckman, 2006, 2007), personality (see Donato et al., 2017), grit (see, e.g., Duckworth et al., 2007; Alan et al., 2019), competitiveness (see, e.g., Buser et al., 2014), economic preferences, and parenting styles (see, e.g., Bonin et al., 2007; Borghans et al., 2008; Cadena and Keys, 2015; Reuben et al., 2017; Falk et al., 2018; Kosse and Tincani, 2020; Falk et al., 2023). While the link between some personal characteristics and

[1] In 2020, 70% of 17-year-olds in Germany were enrolled in some form of upper secondary education; see, e.g., https://stats.oecd.org/Index.aspx?DataSetCode=EAG_ENRL_RATE_AGE#.

(labor market) outcomes is quite well established, others are far less understood. Most importantly, it is not a priori clear if and how the above characteristics interact with specific payments schemes.

Our results show that baseline productivity in the task and one's own assessment of relative performance are the main drivers of output under all three payment schemes. Personality traits, economic preferences and socio-economic background have at best a marginal influence, which basically confirms that incentives do work, by and large, independently of those personal characteristics. This is not the case, however, as far as sorting is concerned. When subjects can choose among the three payment schemes, personality traits, preferences and socio-economic background matter on top of baseline productivity and expectations about own performance. Extraversion and neuroticism, competitiveness, risk and time preferences are predictive of what kind of payment scheme a person chooses. Moreover, sorting does not seem to mainly be focused on optimizing performance. Rather, we can show that – on average – our subjects would benefit both in earnings and utility from the task if an algorithm was applied to assign them to a particular payment scheme, rather than them having the choice.

Our study makes three main contributions. First, we address unanswered questions on heterogeneity in effort provision. While the literature on the interaction between payment schemes and people's characteristics is still scarce, a notable exception is Donato et al. (2017). In the domain of health care provision, they report that people with high conscientiousness (as one of the Big-5 personality traits) provide better maternal and child services, but react less to performance incentivization. People with low conscientiousness and neuroticism perform well with performance incentivization. Moreover, in a lab experiment, Segal (2012) finds a similar pattern (albeit only for men) between conscientiousness and reaction to incentives. Compared to these papers, we present a systematic account of a much larger variety of traits, preferences and socio-demographic characteristics and show in particular how they interact with different payment schemes.

Second, we present an extensive analysis of sorting decisions across three different payment schemes. While the previous literature has usually been limited to studying the sorting decisions between two payment schemes, most prominently between variable payment and tournament incentives (Niederle and Vesterlund, 2007; Buser et al., 2014; Almås et al., 2016), our comprehensive setup allows us to investigate sorting decisions in much greater detail. In addition, we include a wide range of socio-demographics, traits, and preferences that have been found on their own to influence sorting, but we can also examine whether those factors have the same influence both on sorting as well as performance

under a specific payment scheme.

Finally, we contribute by investigating which factors determine performance. Do participants understand their performance potential and how it might change across payment schemes? We find that different characteristics matter in determining performance based on whether one is assigned to or choosing the payment schemes. Having the choice therefore plays an important role in assessing which personal characteristics are important for higher performance. However, participants are not mainly sorting based on the characteristics that are influential when assigned to a payment scheme. This points to other factors being prioritized when choosing a payment scheme. It is the latter aspect that is a key distinction and contribution of our paper in comparison to the most closely related paper that is by Opitz et al. (2024). They ran an experiment on MTurk and studied which personal characteristics were the main drivers for effort in a real effort task under six different, exogenously implemented payment schemes. From this main experiment they can estimate the factors that are related to higher performance, and with these estimates they then let a machine learning algorithm assign a new set of MTurkers to the most promising payment scheme, conditional on the new workers' personality traits. The algorithmic assignment increases performance significantly above the level of the single best payment scheme. While we can also estimate how much algorithmic assignment could improve performance – and also utility – our *Endogenous* treatment allows for three further contributions compared to Opitz et al. (2024). First it reveals additional insights into the drivers of choosing a particular payment scheme. Second it shows that choices of payment schemes are driven by partly different traits than performance under a given payment scheme. Third it shows that subjects fail in maximizing a particular objective (be it utility or performance or earnings) when given the choice between payment schemes.

The remainder of the paper is organized as follows. In section 2 we present our experimental design. Section 3 reports the results. Section 4 provides insights into how an algorithmic assignment to payment schemes could improve performance, earnings and utility of participants. Finally, section 5 concludes the paper.

2 Experimental Design

2.1 Sample

The experiment was conducted with adolescents in schools across North Rhine-Westphalia, Germany (see Appendix A.1 for a map of participating schools). Altogether 1,933 high school students, enrolled in tenth to thirteenth grade, were recruited and attended both sessions. Summary statistics of the students in our study are presented in Table 1 (details

on the variables and measures are explained in the remainder of this section). We targeted what in German is referred to as "Gesamtschule": schools that commonly comprise both low and high education tracks within the same institution. This ensured a heterogeneous sample in our study (with respect to SES, cognitive abilities, etc.). We contacted in a random order the 201 closest schools in the areas of Bonn, Cologne, and Düsseldorf within the state of North Rhine-Westphalia.[2] We first informed and invited schools to participate in the study via a letter. In case of no reply, we contacted the school via phone and sent a more detailed description of the study via email. For every participating school, the study was approved by school principals. Parents were informed about the experiment and needed to sign a consent form in order for a student to participate in the study.[3] Participation was voluntary and it was explicitly mentioned to participants that they could quit the study (or skip specific parts) at any time. As Riener et al. (2020) document the absence of self-selection of schools into experiments in North Rhine-Westphalia, which is where we conducted our experiment, we may assume that our sample is representative of the population of schools in this federal state.

2.2 Real Effort Task (RET)

We implemented a counting task adapted from Abeler et al. (2011). Subjects were presented with a sequence of tables containing zeros and ones (Figure 1). The task consists of highlighting and counting the ones present in each table (for instructions, see Section A.4). A table is correctly solved if: (i) all the ones are highlighted, (ii) none of the zeros are highlighted, and (iii) the total amount of ones is correctly reported. Subjects had a total of three trials to solve each table. The task has several desirable features: it does not require any prior knowledge, performance is easy to objectively measure, learning plays only a minor role, and performing the task has no value outside the experiment (Abeler et al., 2011; Charness et al., 2018). Moreover, the task is tedious and requires effort to be solved.

2.3 Timeline and Treatments

The experiment comprised two parts conducted approximately two weeks apart from each other (see Table 2 for an overview over the two parts).[4] The same subjects participated in both parts of the experiment.[5] In part 1, a broad range of socio-demographic character-

[2]For schools in North Rhine-Westphalia, contact information is publicly available online.

[3]Students that are 18 or older could sign the consent form themselves.

[4]On a few occasions, the two parts were moved closer/further apart due to logistic reasons. Overall, the median time between first and second part of the study was 14 days (Mean (SD) of 12.9 (7.6)). Controlling for the number of days between first and second part of the study leaves our results virtually identical, see Tables A7 and A8 in Appendix A.2.

[5]Schools greatly contributed to obtaining a high rate of pupils participating in both parts by trying to arrange the sessions on the same weekday and the same time, keeping other organizational constraints

Table 1: Summary Statistics

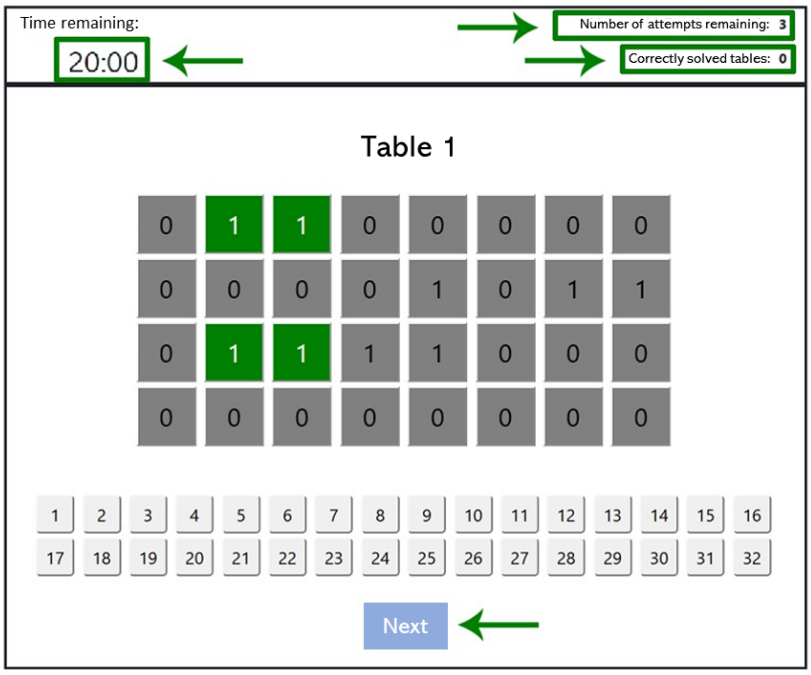
Characteristic	Mean	SD	N
Outcome of Interest			
Performance in part 2 (RET 20 min)	122.24	22.73	1933
Skills			
Grade German	2.73	0.91	1933
Grade Math	2.81	1.16	1933
IQ (Raven 0-10)	5.08	1.46	1933
Productivity part 1 (RET 5 min)	26.94	6.24	1933
Demographics			
Age rel. to grade mean	-0.01	0.81	1933
Female (=1)	0.53	0.49	1933
Grade (9-13)	11.29	1.00	1933
Number of Siblings	1.64	1.12	1933
Positive parenting (cont; 1-5)	3.39	0.91	1933
Composite SES Index	0.00	1.00	1933
Personality Traits, Economic Preferences, and Beliefs			
Altruism (0-10)	7.49	2.22	1933
Extraversion (cont; 1-5)	3.48	0.78	1933
Agreeableness (cont; 1-5)	3.56	0.58	1933
Conscientiousness (cont; 1-5)	3.35	0.40	1933
Neuroticism (cont; 1-5)	2.90	0.72	1933
Openness (cont; 1-5)	3.43	0.65	1933
Enjoy competition (cont; 1-5)	3.11	1.00	1933
Grit (cont; 1-5)	3.23	0.45	1933
Belief on rel. performance (0-1)	0.55	0.25	1933
Overplacement (=1)	0.59	0.49	1933
Patience Index	-0.01	0.80	1933
Risk Taking Index	0.01	0.75	1933

Note: We define all variables in full detail in Section A.3, and present a brief description of the variables in Sections 2.4 and 2.5.

istics, traits, and preferences, as well as a measure of individual baseline productivity was collected from the subject pool. In part 2, effort provision was measured under three different payment schemes: Fixed, Variable, and Tournament payment. Two between-subjects treatments were implemented, in which we either imposed a specific payment scheme (*Exogenous* treatment) or subjects could choose their preferred payment scheme (*Endogenous* treatment). In the remainder of this section, we describe in detail the payments schemes, the treatments, the sequence of tasks in the two parts of the experiment, and the data that we collected.

constant. Nevertheless, this was not always possible, and in addition, participants fell sick between first and second visit or made a mistake in entering their ID (explained below in 'Procedures'). As a result, we were unable to match data from the second part for about 15 percent of first-part participants, which is only slightly higher than the typically to be expected 10 percent sick rate on a single point in time. Importantly, productivity in part 1 does not differ between matched and unmatched first-part participants. Also note that the attrition rate in the classical sense (i.e., dropout rate after treatments have been administered) – which is a threat to internal validity – is zero, as all participants decided to finish the session once started.

Figure 1: Real Effort Task (RET)



Notes: In the center of the screen, participants see a 4×8-table measuring 4×8, where each square can either display a one or a zero. The objective is to click on every square containing the number one to highlight it. Once this is done, participants should provide the total count of highlighted ones by selecting the corresponding number in the lower white cells. To finalize the task, participants need to click on the "Next" button. If all the ones are correctly highlighted, and the accurate count is submitted, a new table will appear. Otherwise, participants have up to three chances to revise their inputs.

Payments Schemes We implemented the following three types of payment schemes:[6]

- Fixed payment: a flat payment (€6.5) independent of the number of correctly solved tables;
- Variable payment: subjects were paid a piece rate (€0.06) per correctly solved table;
- Tournament payment: subjects were paid either a high (€0.08) or a low (€0.04) piece rate per correctly solved table. Each participant in this treatment was matched with another participant that also chose the Tournament payment and was paid the high (low) rate if they solved more (less) tables than the matched participant.[7]

Treatments We ran two between-subjects treatments: *Exogenous* and *Endogenous*. In the *Exogenous* treatment, participants were assigned to either the Fixed, the Variable, or the Tournament payment. Subjects only received information about the relevant payment scheme they were assigned to, and were paid accordingly. In the *Endogenous* treatment, participants received information about all the three types of payment schemes and had to choose one scheme which determined how their payoff was calculated.

[6]Payment amounts for the payment schemes were calibrated based on pilot data to ensure comparable payoffs across payment schemes.

[7]Ties were solved by a random draw of the computer and participants were made aware of that.

Table 2: Timeline and overview of the experimental tasks and design

Part 1 – Socio-demographics, traits and preferences	***Exogenous* treatment**	***Endogenous* treatment**
Personal ID	✓	✓
RET instructions	✓	✓
Productivity (RET 5 min)	✓	✓
IQ test (Raven's matrices, 5min)	✓	✓
SES questionnaire	✓	✓
Big Five (BFI-44)	✓	✓
Competitiveness (14-item)	✓	✓
Economic preferences module	✓	✓
Positive parenting (6-item)	✓	✓
Grit (8-item)	✓	✓
Average payment (€)	€4 + RET (5 mins)	€4 + RET (5 mins)
Average time	45 min	45 min

Part 2 – Effort provision and payment schemes	***Exogenous* treatment**	***Endogenous* treatment**
Personal ID	✓	✓
Belief elicitation	✓	✓
RET instructions	✓	✓
Instructions (payment schemes)	1 payment scheme (within session randomization)[§]	all 3 payment schemes
Choice (payment scheme)	-	✓
RET (20 min)	✓	✓
Average payment (€)	€1 + RET (20 min) + belief	€1 + RET (20 min) + belief
Average time	45 min	45 min
Number of Observations	983	950

Notes: [§]About 1/3 of participants were assigned to each of the three payment schemes.

2.4 Part 1 – Socio-Demographic Characteristics, Traits and Preferences

The first part of the study was common to all treatments and measured a number of characteristics. We focused on three main areas: skills, demographics, as well as preferences, personality traits and beliefs. Here, we provide a short overview over the collected measures, while Appendix A.3 contains a detailed description of all the items as well as the motivation behind the selection of included variables (also see the full questionnaire in Section A.4).

We started by collecting a reference measure of productivity in the real effort task (RET), where we follow Dohmen and Falk (2011) in incentivizing the task. Participants were given five minutes to solve as many tables as they could and were paid on a piece rate (€0.06) basis.[8] From this, we create a residualized productivity measure to be used as predictor for performance in part 2, along all other predictors. We regress, first, our full set of predictors on the number of solved tasks in the five-minute RET (in part 1). Then we

[8]To familiarize themselves with the task and the software, subjects were asked to solve a trial table before moving to the actual task.

use the residuals of this regression as a measure of baseline productivity that is corrected for the correlation between the number of solved tasks in part 1 and the remaining set of our predictors. After the five-minute RET to assess productivity, a five-minute computerized version of a standardized non-verbal intelligence test was administered (matrix task; Raven, 2000).

We then collected demographic information, as well as information about socio-economic status (SES). Our items are informed by three different socio-economic indices, where we also added some own questions (see Appendix A.3 for a detailed list). Given that we elicited fifteen variables to capture socio-economic status, we rely on principal component analysis (PCA) in order to best utilize the extensive data we obtained: Using the weights on the first component resulting from the PCA, we construct a single-item socio-economic status measure that we refer to as composite SES index. It includes all of the items of three different socio-economic status indices, as well as migration indicators and the amount of pocket money that the participants receive, see Appendix A.3 for details.

With respect to psychological measures, we collect the Big Five (John and Srivastava, 1999), competitiveness (Newby and Klein, 2014), positive parenting style (Frick, 1991; Essau et al., 2006), and grit (Duckworth and Quinn, 2009). For all these measures, we rely on widely used psychological scales. Finally, we included a series of non-incentivized questions taken from the validated preference module by Falk et al. (2018, 2023) to measure patience, the willingness to take risk, and altruism.[9]

2.5 Part 2 – Effort Provision and Incentives

The second part of the study captured effort provision under the different payment schemes. The RET was the same as in part 1, but lasted for 20 minutes.[10] Before the RET, we elicited participants' beliefs about the number of correctly solved tasks in the five-minute RET in part 1. More precisely, we ranked all the participants present in the room based on the number of solved tables in part 1 and then asked them to guess their rank.[11] If the guess was correct, they earned €2. If the difference between the guess and the actual ranking was at most 5 positions, they still earned €0.50. Only participants who were present in both visits were included in the ranking (and were asked to guess). We opted for collecting the guesses about their ranking in the second visit. We did this as we wanted to examine how beliefs about rankings relate to the choice of payment scheme and to other variables

[9] For patience and the willingness to take risk, we include both qualitative and quantitative items (see the instructions and the questionnaire in Section A.4), which we aggregate following Falk et al. (2018), i.e., by adding the (weighted) z-scores of qualitative and quantitative items using the weights from the validation.

[10] The distribution of number of solved tables in part 1 and part 2 is shown in Figure A2.

[11] On average there were 34.5 participants per session.

collected in the second session. As students could update their beliefs between the first and the second session, eliciting beliefs at the beginning of the second session ensures that we measure the beliefs that are most relevant for behavior in part 2. Feedback about the guessing task was given only at the end of the study.

The beliefs elicited in this way are used in two measures: 'Belief on rel. performance' is a normalized measure of the subject's belief about their own rank in the distribution of tables solved in the five-minute RET in part 1 of the study. As session sizes differ across observations, we normalize the belief on the own rank by the total number of participants in the respective session. The measure is, thus, defined between 0 (subject believes to be on the lowest percentile in the distribution) and 1 (subject believes to be on the highest percentile in the distribution). 'Overplacement', the second measure, is an indicator for a positive difference between the belief about one's own ranking regarding tasks solved in part 1 and actual own ranking in the five-minute RET in part 1 (both normalized on the session level).

2.6 Assignment to Treatments

The assignment of subjects into treatments happened on the session level. In Table 3, we show that samples assigned to either of the two treatments, *Exogenous* and *Endogenous*, are comparable and do not differ more than expected purely by chance along the elicited characteristics as used in our analyses.[12]

In the *Exogenous* treatment we randomly assigned participants on the individual level to one of the three payment schemes based on the min MSE method developed by Schneider and Schlather (2017).[13] To achieve balanced treatment groups, we consider pre-treatment information on the baseline productivity in the RET, demographic information, socio-economic characteristics, psychological measures as well as preferences. In Table 4, we investigate whether our explanatory variables across different payment schemes in *Exogenous* are balanced by testing whether at least one group is different from the other two groups using Kruskal-Wallis tests. Of the 22 comparisons, none is significant not even at the 10% level, indicating that our treatment assignment was successful in creating well balanced groups.

[12] In particular, by chance, one would expect 10 percent of the 22 variables to differ significantly at the 10% significance level, that is, for two variables, we would expect such a difference by chance. Indeed, two variables differ to such a degree: IQ, and the number of siblings. The differences are, however, small compared to the scale and the SD (reported in parentheses).

[13] Based on re-randomization, this method aims at minimizing the mean squared error of the treatment effect estimator as a function of treatment assignment. The method thus increases precision of the treatment effect estimation by choice of treatment assignment. Intuitively, the method forms comparable treatment groups considering multivariate information such as baseline productivity, gender, SES, etc. We opt for this method, as it allows us to assign three treatments in the same session while still "balancing" multivariate and continuous information in a principled way (previous work mostly relies on binary assignment between two different treatments).

In the *Endogenous* treatment, students were first introduced to the three available payment schemes and were then allowed to pick their most preferred one before starting to work for 20 minutes on the RET.

Table 3: Balance in treatment assignment for *Exogenous* and *Endogenous*

		Treatments		
Characteristic	**Overall**	**Exogenous**	**Endogenous**	**p-value**
Skills				
Grade German	2.73 (0.91)	2.72 (0.92)	2.74 (0.91)	0.77
Grade Math	2.81 (1.16)	2.82 (1.13)	2.79 (1.19)	0.50
IQ (Raven 0-10)	5.08 (1.46)	5.13 (1.43)	5.02 (1.49)	0.09
Productivity part 1 (RET 5 min)	26.94 (6.24)	26.84 (6.19)	27.03 (6.29)	0.53
Demographics				
Age rel. to grade mean	-0.01 (0.81)	0.01 (0.77)	-0.03 (0.84)	0.11
Female (=1)	0.53 (0.49)	0.53 (0.49)	0.54 (0.49)	0.72
Grade (9-13)	11.29 (1.00)	11.28 (0.97)	11.29 (1.02)	0.91
Number of siblings	1.64 (1.12)	1.68 (1.13)	1.59 (1.11)	0.07
Positive parenting (1-5)	3.39 (0.91)	3.40 (0.91)	3.38 (0.91)	0.49
Composite SES Index	0.00 (1.00)	-0.03 (0.99)	0.03 (1.01)	0.25
Personality Traits, Economic Preferences, and Beliefs				
Altruism (0-10)	7.49 (2.22)	7.53 (2.20)	7.46 (2.25)	0.57
Extraversion (1-5)	3.48 (0.78)	3.47 (0.78)	3.48 (0.78)	0.95
Agreeableness (1-5)	3.56 (0.58)	3.56 (0.57)	3.57 (0.59)	0.46
Conscientiousness (1-5)	3.35 (0.40)	3.36 (0.42)	3.35 (0.39)	0.66
Neuroticism (1-5)	2.90 (0.72)	2.91 (0.73)	2.88 (0.71)	0.42
Openness (1-5)	3.43 (0.65)	3.41 (0.66)	3.45 (0.65)	0.24
Enjoy competition (1-5)	3.11 (1.00)	3.09 (0.98)	3.12 (1.01)	0.56
Grit (1-5)	3.23 (0.45)	3.23 (0.46)	3.22 (0.45)	0.41
Belief on rel. performance (0-1)	0.55 (0.25)	0.55 (0.24)	0.54 (0.25)	0.35
Overplacement (=1)	0.59 (0.49)	0.60 (0.49)	0.57 (0.49)	0.25
Patience Index	-0.01 (0.80)	0.00 (0.79)	-0.01 (0.82)	0.94
Risk Index	0.01 (0.75)	0.00 (0.75)	0.02 (0.76)	0.53
Number of Observations	**1933**	**983**	**950**	

Note: The p-values report results from Wilcoxon rank-sum tests of differences between the two treatment groups. Standard deviations in parentheses.

2.7 Procedures

To obtain a heterogeneous sample in terms of background characteristics, to limit experience with different payment schemes and to avoid self-selection into the study, we conducted our study in schools during regular school hours. Sessions were run in large lecture halls and several classes took part in the experiment at the same time.[14] The number of participants in a single session was on average 34.5 with a 12.9 standard deviation. The experiment was conducted with up to 75 tablets and a server using oTree (Chen et al., 2016). We aimed for a sample size of about 2000, as a result of analytical and simulation-

[14]Due to logistic constraints, some sessions were conducted in single classrooms.

Table 4: Balance for Payments in *Exogenous*

		Incentive Schemes			
Characteristic	**Overall**	**Fixed**	**Variable**	**Tournament**	**p-value**
Skills					
Grade German	2.72 (0.92)	2.71 (0.96)	2.72 (0.88)	2.72 (0.91)	0.94
Grade Math	2.82 (1.13)	2.81 (1.07)	2.83 (1.18)	2.82 (1.13)	0.99
IQ (Raven 0-10)	5.13 (1.43)	5.14 (1.37)	5.16 (1.45)	5.09 (1.49)	0.84
Productivity part 1 (RET 5 min)	26.84 (6.19)	26.91 (6.25)	26.77 (6.52)	26.83 (5.79)	0.91
Demographics					
Age rel. to grade mean	0.01 (0.77)	0.03 (0.78)	-0.04 (0.75)	0.02 (0.78)	0.48
Female (=1)	0.53 (0.49)	0.52 (0.50)	0.53 (0.49)	0.53 (0.49)	0.94
Grade (9-13)	11.28 (0.97)	11.32 (0.98)	11.27 (0.95)	11.25 (0.99)	0.60
Number of siblings	1.68 (1.13)	1.62 (1.09)	1.71 (1.12)	1.73 (1.17)	0.49
Positive parenting (1-5)	3.40 (0.91)	3.38 (0.91)	3.40 (0.91)	3.42 (0.92)	0.79
Composite SES Index	-0.03 (0.99)	0.01 (1.03)	-0.05 (0.93)	-0.04 (1.01)	0.74
Personality Traits, Economic Preferences, and Beliefs					
Altruism (0-10)	7.53 (2.20)	7.47 (2.27)	7.54 (2.20)	7.57 (2.12)	0.94
Extraversion (1-5)	3.47 (0.78)	3.47 (0.77)	3.46 (0.78)	3.50 (0.80)	0.80
Agreeableness (1-5)	3.56 (0.57)	3.54 (0.56)	3.55 (0.57)	3.58 (0.57)	0.85
Conscientiousness (1-5)	3.36 (0.42)	3.34 (0.42)	3.35 (0.40)	3.37 (0.44)	0.76
Neuroticism (1-5)	2.91 (0.73)	2.95 (0.76)	2.92 (0.70)	2.88 (0.73)	0.56
Openness (1-5)	3.41 (0.66)	3.39 (0.68)	3.44 (0.60)	3.40 (0.68)	0.62
Enjoy competition (1-5)	3.09 (0.98)	3.08 (1.00)	3.13 (0.95)	3.07 (0.99)	0.73
Grit (1-5)	3.23 (0.46)	3.23 (0.50)	3.24 (0.46)	3.24 (0.42)	0.81
Belief on rel. performance (0-1)	0.55 (0.24)	0.54 (0.25)	0.56 (0.24)	0.56 (0.24)	0.31
Overplacement	0.60 (0.49)	0.60 (0.49)	0.59 (0.49)	0.60 (0.49)	0.94
Patience Index	0.00 (0.79)	0.04 (0.78)	-0.02 (0.80)	-0.02 (0.78)	0.65
Risk Index	0.00 (0.75)	0.00 (0.77)	0.00 (0.73)	0.00 (0.74)	0.90
Number of Observations	**983**	**331**	**327**	**325**	

Note: The p-values report results from Kruskal-Wallis tests of whether at least one group is different from the other groups. Standard deviations in parentheses.

based power calculations.[15] Data was collected between March 2019 and August 2022.[16]

In each of the two parts, subjects were randomly assigned to a desk upon arrival. They were all separated by privacy screens, and communication was strictly forbidden throughout the experiment. This was enforced to avoid students comparing choices or their performance. Teachers were allowed to be in the classroom but were not allowed to communicate with or observe the behavior of the participants. In the first part, the relevant instructions were read aloud, and displayed on the screens before the beginning of the RET and IQ task. In the second part, subjects were reading the instructions displayed on the screen alone, since multiple payment schemes were randomized within the same experimental session in the *Exogenous* treatment. To ensure that subjects fully understood the

[15]With this sample size, we have 80% power for detecting interaction effects of moderate size (i.e., 30% of a main effect), see our pre-registration for details and formulae.

[16]Data collection was paused several times during the COVID-19 pandemic because of school closures. Although we had pre-registered to end data collection in May 2022, for some schools, it was not possible to offer us a date for conducting sessions before, even though we had contacted them at the beginning of the school year, i.e., September of the previous year. We accepted their offer for later dates nonetheless, thereby maintaining a good relationship with schools and openness towards future research requests by us and others in our region.

payment schemes, they had to individually answer a set of computerized control questions before proceeding with the task itself.

Since subjects participated in two separate parts, data was matched via a personal ID created by the participants at the beginning of each part (see the instructions in Section A.4). Each part lasted around 45 minutes (a regular school hour) and participants were paid anonymously and in cash. In part 1, participants received a fixed payment of €4, plus the earnings for the 5 minutes RET. In part 2, participants earned a €1 show-up fee, plus the earnings for the 20 minutes RET and from the guessing task (beliefs). On average, participants earned €5.65 in part 1 and €8.71 in part 2, which is in total roughly in the range of what is recommended as weekly allowance for that age group.[17]

3 Results

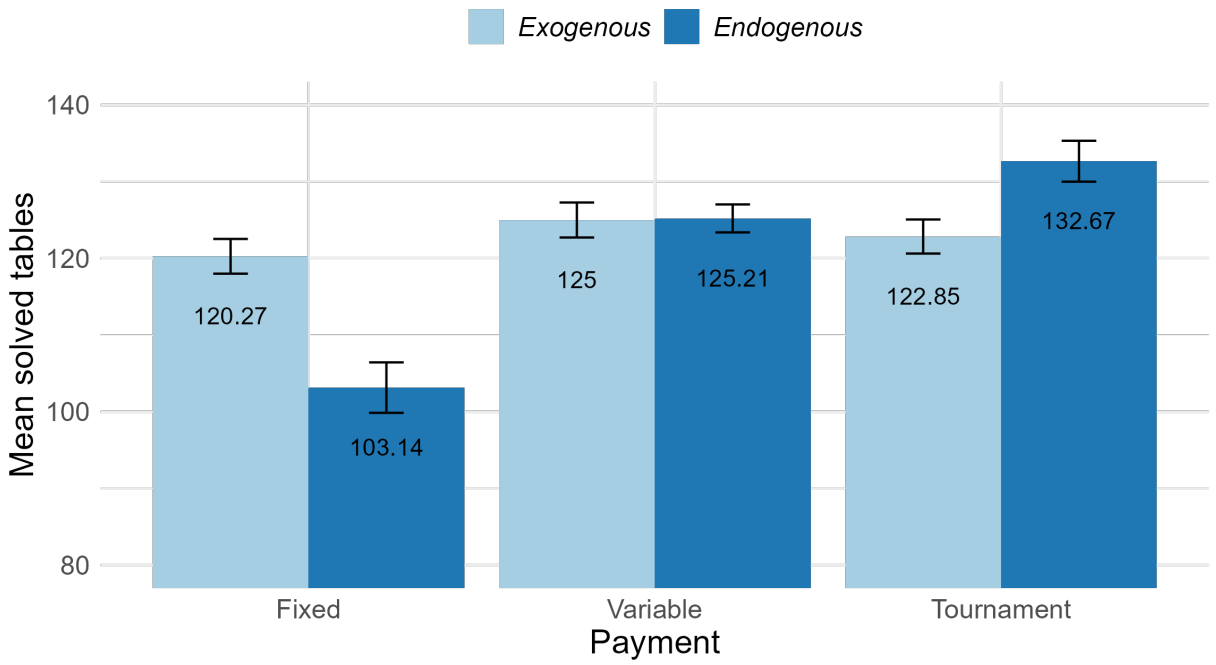
3.1 Heterogeneity in Effort Provision – *Exogenous* treatment

The light blue bars in Figure 2 illustrate performance across payment schemes in the *Exogenous* treatment. Performance is measured as the total number of correctly solved tables in the 20-minute real effort task. We see that average performance is in the range from 120 to 125 correct tasks. Despite the relatively small range, we see significant differences across payment schemes, as participants with Variable payment performed, on average, better than participants with Fixed payment (t-test, p <0.01) and those under Tournament payment (p <0.05).

Next we investigate this heterogeneity in effort provision by examining how they are related to individual characteristics. The results are reported in Table 5. We have structured the table into three groups of variables, with skills at the top, demographics in the middle and personality traits, economic preferences and beliefs at the bottom. In the upper panel, we see that math grades and baseline productivity (measured as the residualized performance in the 5-minute RET in part 1 of the experiment) have a positive influence on performance (in part 2). IQ is unrelated to performance. In the middle panel, we note that participants in higher grades – i.e., older participants – (recall they are between 16 and 20 years old) are better performing, while those who are relatively older in their respective grade perform worse. The latter might likely be due to participants who had to repeat a grade (which happens in about 10% to 20% of cases in Germany) and are thus relatively older (and given the repetition of a grade on average less able, motivated, or focused than others). Interestingly, socio-economic status is not related to performance. In the bottom panel, it is noteworthy that not a single personality trait of the Big-5 has

[17] See, e.g., https://www.dji.de/themen/jugend/taschengeld.html (in German; last accessed: 12/07/2023).

Figure 2: Performance Measured by the Average of Correctly Completed Tasks in Part 2 Conditional on Treatment and Payment Scheme



Notes: Whiskers (in black) indicate 95% confidence intervals. In the *Exogenous* treatment, there is a significant difference in performance, where those participants assigned to Variable payment performed better than both, those assigned to Fixed payment as well as those assigned to Tournament payment (paired t-test of difference: $p < 0.01$ and $p < 0.05$, respectively). There are no differences between Fixed and Tournament payment. In the *Endogenous* treatment, we observe a significant difference in performance in all pairwise comparisons of payment conditions, where those with Fixed payment performed the poorest, followed by those with Variable payment, and participants under Tournament payment performing best (paired t-test in all pairwise comparisons: $p < 0.01$).

a significant influence on performance, nor are economic preferences (risk and time preferences, competitiveness) or grit relevant. Only one's belief about the relative productivity (measured based on the number of tasks solved in the five-minute RET in part 1) as well as overplacement is significant. The first variable, relative productivity, is scaled from 0 to 1, spanning the range from expecting to be the worst performer (0) to expecting to be the best performer (1). The coefficients of around 30 indicate that someone is estimated to solve 30 more tasks if the person believes to be the best performer rather than the worst performer. In other words, if someone believes to move up one decile in relative performance, the person can be expected to solve about 3 tasks more. The coefficient of overplacement shows that those who think that they have been more productive in the five-minute RET in part 1 than they actually were, solve, on average, between about 14 and 16 tables less.[18]

[18]Having more optimistic and having overoptimistic beliefs (i.e., overplacement) is correlated mainly with positive parenting, and negatively correlated with being female (see Table A5).

Overall, it seems that there are hardly any differences regarding the predictive quality of the individual characteristics across payment schemes, and a single regression with interaction terms confirms this (see Table A1 and Table A2 in Appendix A.2). The most sizeable difference worth highlighting concerns the math grade (highest and positive correlation for Fixed, lowest, insignificantly positive correlation in the case of Tournament payment). More generally, and irrespective of differences between the payment schemes, the results in Table 5 suggest that personality traits and SES are not very influential for performance under the different payments schemes when they are imposed exogenously.

Table 5: Productivity by Payment Scheme – *Exogenous* treatment

	Payment Schemes		
	Fixed	**Variable**	**Tournament**
	(I)	(II)	(III)
Skills			
Grade German	-1.113 (0.927)	-0.758 (1.782)	-0.063 (1.129)
Grade Math	4.274 (0.893)***	2.070 (0.856)**	1.112 (0.914)
IQ (Raven 0-10)	0.162 (0.525)	0.356 (0.633)	0.944 (0.588)
Productivity (resid.)	1.573 (0.270)***	1.400 (0.355)***	1.712 (0.407)***
Demographics			
Age rel. to grade mean	-2.189 (0.760)***	-3.142 (1.288)**	-1.515 (1.282)
Female (=1)	4.143 (2.627)	1.881 (2.573)	0.806 (2.378)
Grade (9-13)	4.112 (0.794)***	3.605 (1.557)**	4.231 (1.040)***
Number of siblings	-0.260 (1.304)	-1.043 (0.829)	0.972 (0.792)
Positive Parenting (1-5)	1.476 (1.354)	-0.302 (1.065)	-1.616 (1.115)
Composite SES Index	-0.122 (0.866)	-0.626 (1.026)	-0.883 (1.161)
Personality Traits, Economic Preferences, and Beliefs			
Altruism (0-10)	-0.392 (0.424)	0.637 (0.478)	-0.072 (0.583)
Extraversion (1-5)	-1.220 (1.311)	0.376 (1.244)	-0.254 (1.550)
Agreeableness (1-5)	1.662 (2.849)	-1.058 (1.976)	-1.604 (1.843)
Conscientiousness (1-5)	0.037 (2.315)	0.813 (4.119)	1.808 (2.825)
Neuroticism (1-5)	-0.250 (1.279)	-0.487 (1.323)	-1.083 (1.555)
Openness (1-5)	0.940 (1.527)	-0.770 (1.917)	0.406 (1.429)
Enjoy Competition (1-5)	-0.392 (1.317)	1.293 (1.259)	1.446 (1.319)
Grit (1-5)	-1.821 (1.864)	-0.638 (2.416)	-1.367 (2.085)
Belief on rel. performance (0-1)	30.142 (4.478)***	30.923 (4.024)***	30.193 (3.884)***
Overplacement (=1)	-14.325 (2.170)***	-13.532 (2.430)***	-15.674 (1.819)***
Patience Index	0.953 (1.103)	0.357 (1.205)	-0.658 (1.520)
Risk Taking Index	1.546 (1.140)	-1.114 (1.159)	-2.136 (1.690)
Constant	50.961 (20.652)**	66.835 (26.034)**	64.905 (19.851)***
Num.Obs.	331	327	325
R2 Adj.	0.343	0.342	0.341

Note: Table shows OLS regressions of performance on characteristics in part 2 split by treatment and payment scheme. Standard errors (in parentheses) clustered on the session level. Productivity (resid.) is a residualized measure of performance in part 1 (see previous section for details).

3.2 Heterogeneity in Effort Provision – *Endogenous* Treatment

The dark blue bars in Figure 2 show large differences in the average performance when payment schemes have been chosen by participants themselves, with the output under Tournament payment being almost 30% higher than under Fixed payment. Comparing the light blue bars (for the *Exogenous* treatment) with the dark blue bars (for the *Endogenous* treatment), we note that self-selection matters a lot. In Fixed, performance drops markedly by about 20% compared to the *Exogenous* treatment, while in Tournament it increases by about 8% in the *Endogenous* treatment. This already hints at selection effects, and we are going to study the factors for selection in the next subsection.

In Table 6 we present the regression results on which factors are related to performance with a given payment scheme. This table is identically structured as the previous Table 5. Comparing both tables, we note that also in the *Endogenous* treatment, three variables are robustly related to performance, which are (the residualized) productivity in part 1 (although not in Fixed), beliefs about one's relative performance, and overplacement. While here the patterns are very similar across both treatments, Table 6 also reveals differences. Age (both absolute as the grade one is in, and relative compared to the grade mean) becomes far less important, and only in one case the coefficient remains significantly different from 0. The bottom panel reveals that personality traits and economic preferences become more important in the *Endogenous* treatment, which may not be so surprising, given that participants can make their own choice about the payment scheme, which is related to personal characteristics, as we will see later. Regarding personality traits, we observe that extraversion is negatively correlated with performance under Fixed payment, and positively under Tournament payment (at the 10% level), with the differences being significant at the 5% level. Similarly, agreeableness is negatively correlated with performance under Fixed payment, although insignificantly so, where the relation flips for Variable and Tournament payment, such that the correlations are significantly positive. In turn, higher conscientiousness is associated with worse performance when choosing Tournament payment, while the reverse is true (albeit insignificant) when choosing Variable or Fixed payment. In line with Donato et al. (2017), we find a negative interaction between conscientiousness and Tournament payment, i.e., a significant difference in coefficients of conscientiousness in Tournament compared to both, Variable and Fixed payment (see Tables A3 and A4 in Appendix A.2).

Finally, there are two more noteworthy results that may seem puzzling at first. One concerns the negative coefficients of grit in Variable, which turns (insignificantly) positive in Tournament (with the difference between the coefficients being significant, see Table A4). As grit is defined as (intrinsic) perseverance toward a set goal (Alan et al., 2019), we might expect in particular the "ungritty" to profit from extrinsic motivation in the form of constant incentivization due to Variable payment. In fact, this is what we observe: the

negative coefficient in Variable is driven by below-gritty individuals, which outweighs the positive coefficient of the above-gritty individuals (that seem to perform well in any case). The second perhaps puzzling result regards the coefficient of willingness to take risk, which is negative in Tournament. Here, it is important to stress that the sample in Tournament is self-selected, and we may expect the risk loving to be over-represented here (more on that below). Keeping ability constant, this results in a negative coefficient of the willigness to take risk.

Table 6: Productivity by incentive scheme – *Endogenous* treatment

	Incentive Schemes		
	Fixed	**Variable**	**Tournament**
	(I)	(II)	(III)
Skills			
Grade German	-4.443 (1.560)***	-1.362 (0.949)	1.035 (1.228)
Grade Math	2.018 (1.274)	1.738 (0.765)**	0.838 (1.090)
IQ (Raven 0-10)	1.132 (1.405)	0.830 (0.503)*	1.327 (0.880)
Productivity (resid.)	0.786 (0.762)	2.252 (0.315)***	2.315 (0.385)***
Demographics			
Age rel. to grade mean	1.420 (1.455)	0.081 (0.973)	-0.057 (1.142)
Female (=1)	0.247 (5.580)	3.370 (1.542)**	2.394 (3.497)
Grade (9-13)	1.084 (1.959)	1.665 (0.822)**	2.257 (1.521)
Number of siblings	-1.540 (1.073)	0.067 (0.723)	0.670 (1.394)
Positive Parenting (1-5)	-1.816 (2.137)	-0.254 (0.662)	0.399 (1.108)
Composite SES Index	0.541 (1.394)	-0.813 (0.491)*	1.673 (1.184)
Personality Traits, Economic Preferences, and Beliefs			
Altruism (0-10)	0.845 (1.054)	0.149 (0.360)	-0.660 (0.412)
Extraversion (1-5)	-4.319 (2.318)*	0.532 (1.391)	2.179 (1.313)*
Agreeableness (1-5)	-3.319 (4.154)	3.141 (1.539)**	2.605 (1.506)*
Conscientiousness (1-5)	6.459 (5.510)	2.915 (2.293)	-6.637 (3.268)**
Neuroticism (1-5)	-1.422 (3.411)	0.698 (1.051)	-1.162 (1.740)
Openness (1-5)	0.485 (2.668)	-2.367 (1.331)*	-0.689 (1.358)
Enjoy Competition (1-5)	-1.160 (1.452)	0.731 (0.695)	-0.851 (1.345)
Grit (1-5)	-1.539 (4.851)	-3.369 (1.475)**	2.931 (2.204)
Belief on rel. performance (0-1)	34.619 (7.656)***	29.954 (3.874)***	27.078 (3.507)***
Overplacement (=1)	-16.678 (3.051)***	-20.741 (1.351)***	-16.807 (2.106)***
Patience Index	-2.908 (2.229)	-0.144 (0.768)	0.429 (1.250)
Risk Taking Index	-0.243 (2.732)	0.283 (0.785)	-2.428 (1.185)**
Constant	108.566 (21.975)***	85.643 (14.452)***	88.358 (23.538)***
Num.Obs.	235	458	257
R2 Adj.	0.134	0.478	0.472

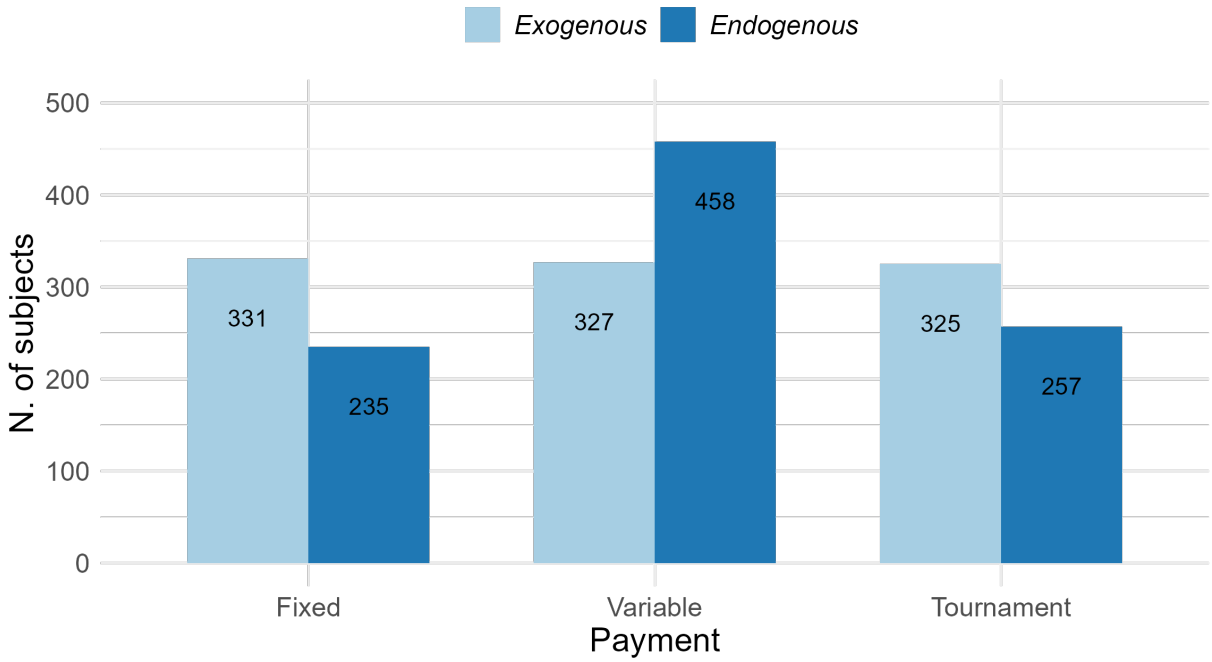
Note: Table shows OLS regressions of performance on characteristics in part 2 split by treatment and incentive scheme. Standard errors (in parentheses) clustered on the session level. Productivity (resid.) is a residualized measure of performance in part 1. For this, we regress our full set of predictors on performance. We use the residuals of this regression as a measure of productivity that is corrected for the correlation between performance in Part 1 and the remaining set of our predictors.

Overall, compared to the *Exogenous* treatment, the evidence from the *Endogenous* treatment suggests that performance in case of having agency over the payment scheme is partly driven by other factors than when the payment scheme is exogenously assigned.

3.3 Determinants of Sorting across Payment Schemes

Figure 3 shows the number of participants that chose each of the payment schemes. The light blue bars refer to the *Exogenous* treatment where the assignment was random, yielding practically the same number of observations for each payment scheme. The dark blue bars for the *Endogenous* treatment reveal that sorting is not random, however. The Variable payment is chosen most often (about half of the time), with the other two payment schemes being roughly similarly attractive and accounting for about a quarter of choices each. Figure 4 shows the average productivity in the 5-minute task in part 1 of the experiment conditional on the selected or assigned payment scheme. Sorting is obviously related to productivity, as the dark blue bars in this figure reveal. Subjects who solved more tables in part 1 are most likely to sort into Tournament payment and least likely to sort into Fixed payment. The light blue bars for the *Exogenous* treatment indicate that performance in part 1 is orthogonal to the random assignment to payment schemes in part 2.

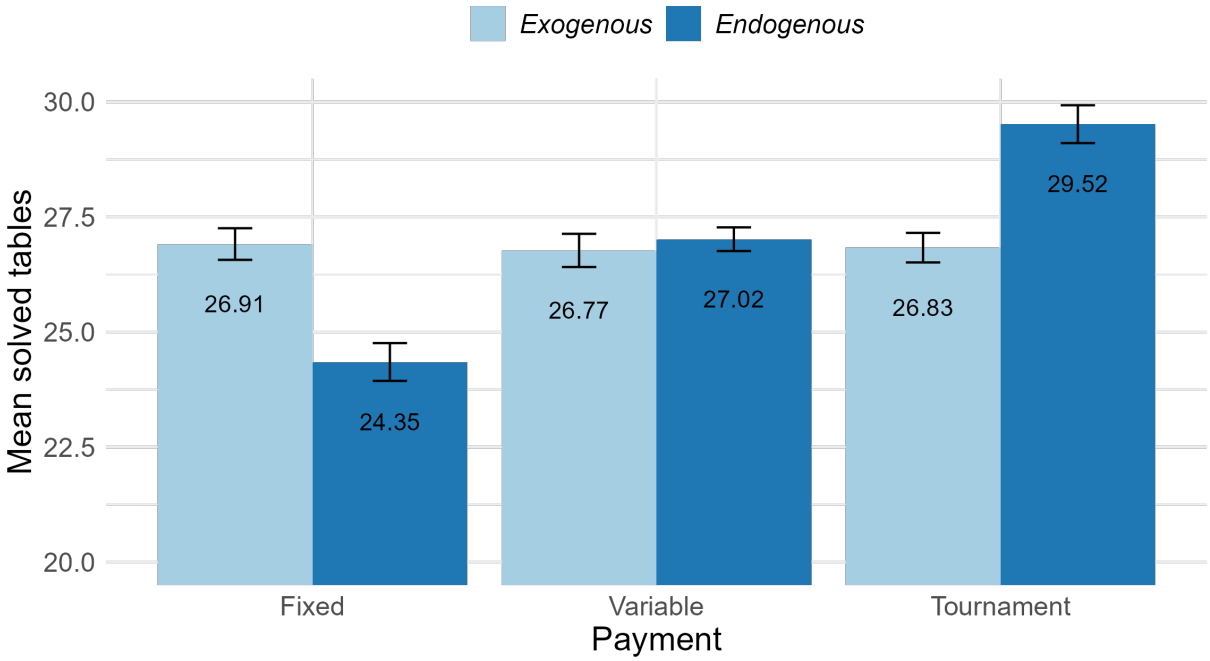
Figure 3: Choices of Payment Schemes: Part 2



Notes: In the *Endogenous* treatment, a significantly higher number of participants choose Variable over the other two payment schemes (paired t-test for each difference: p <0.01). The number of participants selecting into Fixed payment does not differ from the number of participants selecting Tournament payment.

We now continue to analyze the influence of skills, demographic and socio-economic information as well as traits, preferences and beliefs on this non-random sorting in the *Endogenous* treatment. Owing to our extensive set of covariates, which is much larger than that of prior studies, we can simultaneously control for all these variables (called 'full model' below). Comparing both the 'full model' and a simple correlation helps to assess the robustness of our results and shed further light on results previously reported in the

Figure 4: Tasks Completed in Part 1, by Treatment and Payment Scheme in Part 2



Notes: In the *Endogenous* treatment, we note that participants choosing the Fixed payment in part 2 performed the poorest in part 1, followed by the ones choosing the Variable payment, with those choosing Tournament payment performing best (paired t-test for each difference: p <0.01). In the *Exogenous* treatment, there are no such differences in part 1.

literature.

We provide the resulting overview regarding sorting decisions in Table 7 (and report the full logit regressions including coefficients in Table A6). Panel A relates skills to sorting, Panel B demographics, and Panel C personality traits, economic preferences and beliefs. In each panel, the first column lists the respective variables that we consider. The middle column then specifies whether we show our own findings (either in a full model with all variables listed in Table 7 or by only reporting correlations between sorting and the respective variable) or whether we refer to findings in the previous literature. The latter means that Table 7 compares how our findings relate to the ones reported in previous papers that have examined sorting and how it relates to the various variables. None of the papers that we refer to in Table 7 have such a broad range of variables as we have, however. And moreover almost all of them have only pairwise comparisons between two payment schemes (rather than between three as in our case). After the middle column in Table 7 we then show the results on the right hand side of the table. The column "Consistent?" indicates whether our findings are in line with previous findings (✓) or not (✗). In the following columns, we show the direction of relationships between sorting and all variables, whether they are significantly positive (∧), significantly negative (∨), or insignificant (○). This is done for all possible pairwise comparisons (with F for Fixed payment, V for Variable

payment and T for Tournament payment), and in the ultimate column also for an ordered logit model (where the order comes from the riskiness of the outcome, from no risk in the Fixed payment over risk that one has control over in the Variable payment to risk that may even be outside one's own influence in the Tournament payment).

Looking at Panel A, we note first of all that our results are almost always in line with findings of earlier papers. The main insight from Panel A is that productivity is essential for sorting (out of Fixed payment and into Variable or Tournament payment), which is not surprising.[19] However, the null-findings for IQ came more as a surprise to us, and is the only noticeable deviation from previous literature (Buser et al., 2014). Given that they had only two payment schemes and proxy IQ by the GPA, however, it is not clear what Buser et al. (2014) would have found with three payment schemes as well, and a more direct measure of IQ (also in light of the different sign they report for the math grade). By and large, we also find that better grades in German and math lead to sorting out of the Fixed payment (either in favor of the Variable or the Tournament payment).

Panel B shows that gender is important for sorting, as the large majority of previous papers (albeit with only two payment schemes) has also found. Yet, our results are much more nuanced than what is commonly reported in the literature. Women have been reported to be less likely to sort into competitive payment schemes (Niederle and Vesterlund, 2007; Buser et al., 2014, 2017). Most of the literature has focused on sorting decisions between a variable payment and a tournament payment (Gupta et al., 2011; Niederle and Vesterlund, 2007; Buser et al., 2014; Almås et al., 2016; Buser et al., 2017; Reuben et al., 2017; Buser et al., 2024, 2021). Here the relationship is unambiguous. Women shy away from tournament payments more often than men if a variable payment is the alternative. This is also what we find. Dohmen and Falk (2011) find no relation between gender and sorting into a tournament payment over a fixed payment or (separately) between sorting into a variable payment over a fixed payment. With our comparison of three payment schemes, we observe that there is a strong tendency of women to self-select into our Variable payment condition compared to the Fixed and Tournament payment (p <0.01). This finding is obtained from pairwise correlations as well as partial correlations adjusting for all other predictors, such as risk aversion, and we see this result irrespective of pooling the Fixed payment with the Tournament payment, or only comparing the Variable payment with the Fixed payment. This indicates that by no means women shy away from performance-based payments per se. The other variables captured in Panel B of Table 7 seem unrelated to sorting.

[19]Note again that we do not include our baseline measure of productivity from part 1 as a predictor as it might cover up the potential explanatory power of other predictors. Instead, we use a residualized measure, as explained in Section 2.4.

Panel C shows results for personality traits, economic preferences and beliefs. Here again our results confirm almost always previous findings, but at the same time extend them by our choice between three different payment schemes (and by the much more encompassing set of explanatory variables). From this part of the table, it becomes clear that the Big Five personality traits matter for sorting (but recall that they hardly mattered for performance in the *Exogenous* treatment). Extraversion and neuroticism are predictive for sorting out of Variable payment into Fixed payment (p <0.05). Competitiveness is also an important predictor for sorting, as a higher score in the Competitive Orientation Measure (Newby and Klein, 2014) is related to a higher likelihood to choose Variable, as well as Tournament payment, and for avoiding a Fixed payment (p <0.01). Beliefs about one's own relative performance also matter for sorting. Individuals who perceive their own productivity to be on the upper end of the distribution are more likely to sort into a Tournament payment compared to both, the Fixed or the Variable payment. Also this finding is consistent with prior findings in, e.g., Dohmen et al. (2011). In addition, and contrary to the null-finding in Dohmen et al. (2011), we find that those with more positive beliefs are also more likely to select into the Tournament payment compared to the Variable payment. For overplacement, we observe the contrary: a higher likelihood to select into the Fixed compared to both, Variable and Tournament payment. This pattern would be consistent with overoptimistic individuals trying to protect their ego (particularly in front of others) by making their payment uninformative about their performance (Castagnetti and Schmacker, 2022). Similarly, we also see – as practically all previous literature – that more risk taking individuals are more likely to sort into Tournament payment, while they do not seem to matter for the preference between the Variable and Fixed payment.

Finally, both grit and altruism have not been studied in the literature so far with respect to their influence on sorting between the payment schemes. While we have seen correlations with performance, for neither of the two we see a robust relation to sorting, at least once we control for all other variables that we have collected.

Table 7: Predictors of sorting decisions

Variables	**Literature**	**Consistent?**	**F ≺ V or T**	**V ≺ F or T**	**V ≺ F**	**T ≺ V or F**	**T ≺ V**	**T ≺ F**	**Ordered logit**
Panel A: Skills									
Grade German	**Our findings - Full model**		∨	○	∧	○	○	○	∧
	Our findings - Correlation		∨	∧	∧	○	○	∧	
Grade Math	**Our findings - Full model**		○	○	○	∧	∧	∧	∧
	Our findings - Correlation		∨	○	○	∧	∧	∧	
	Buser et al. (2014)	✓					∧		
IQ	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	

To be continued on next page ...

Variables	Our Findings/Literature	Consistent?	F ⋏ V or T	V ⋏ F or T	V ⋏ F	T ⋏ V or F	T ⋏ V	T ⋏ F	Ordered logit
	Buser et al. (2014)†	(✗)					∨		
Productivity (resid.)	**Our findings - Full model**		∨	○	∧	∧	∧	∧	∧
	Our findings - Correlation		∨	○	∧	∧	∧	∧	
	Almås et al. (2016)	✓					∧		
	Dohmen and Falk (2011)§	✓	∨		∧			∧	
	Fornwagner et al. (2023)	✓					∧		
	Niederle and Vesterlund (2007)	✓					∧		
	Reuben et al. (2017)	(✗)					○		
	Sutter and Glätzle-Rützler (2015)	✓					∧		
	Tungodden and Willén (2023)	✓					∧		
Panel B: Demographics									
Age rel. to grade mean	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Female	**Our findings - Full model**		∨	∧	∧	∨	∨	○	○
	Our findings - Correlation		○	∧	∧	∨	∨	○	
	Almås et al. (2016)	✓					∨		
	Boneva et al. (2022)	✓					∨		
	Buser et al. (2014)	✓					∨		
	Buser et al. (2017)	✓					∨		
	Buser et al. (2024)¶	✓					∨		
	Buser et al. (2021)	✓					∨		
	Buser et al. (2022)	✓					∨		
	Gupta et al. (2011)	✓					∨		
	Dohmen and Falk (2011)§	(✗)	○		○			○	
	Eriksson et al. (2009)	(✗)					○		
	Fornwagner et al. (2023)	✓					∨		
	Niederle and Vesterlund (2007)	✓					∨		
	Reuben et al. (2017)	✓					∨		
	Sutter and Glätzle-Rützler (2015)	✓					∨		
	Tungodden and Willén (2023)	✓					∨		
Grade	**Our findings - Full model**		○	∨	○	○	∧	○	○
	Our findings - Correlation		○	∨	○	○	∧	○	
	Sutter and Glätzle-Rützler (2015)	(✗)					○		
Number of siblings	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Positive parenting	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Socio-economic status	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
	Almås et al. (2016)	(✗)					∧		
	Boneva et al. (2022)$	✓					○		
	Tungodden and Willén (2023)	(✗)					∨		
Panel C: Personality Traits, Economic Preferences and Beliefs									
Altruism	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	∧	○	∨	∨	∨	

To be continued on next page ...

Variables	Our Findings/Literature	Consistent?	F ⋏ V or T	V ⋏ F or T	V ⋏ F	T ⋏ V or F	T ⋏ V	T ⋏ F	Ordered logit
Big 5	**Our findings - Full model**								
	Extraversion		∧	∨	∨	○	○	○	○
	Agreeableness		○	○	○	○	○	○	○
	Conscientiousness		○	○	○	○	○	○	○
	Neuroticism		∧	∨	∨	○	○	○	○
	Openness		○	○	○	○	○	○	○
	Our findings - Correlation								
	Extraversion		○	∨	○	○	∧	○	
	Agreeableness		○	∧	○	∨	∨	∨	
	Conscientiousness		○	∧	○	○	○	○	
	Neuroticism		∧	○	○	∨	∨	∨	
	Openness		○	○	○	○	○	○	
	Almås et al. (2016)	✓							
	Extraversion						○		
	Agreeableness						○		
	Conscientiousness						○		
	Neuroticism						○		
	Openness						○		
Enjoy Competition	**Our findings - Full model**		∨	○	∧	∧	∧	∧	∧
	Our findings - Correlation		∨	∨	○	∧	∧	∧	
	Buser et al. (2017)	✓					∧		
	Niederle and Vesterlund (2007)	✓					∧		
	Reuben et al. (2017)	✓					∧		
Grit	**Our findings - Full model**		○	○	○	○	○	○	○
	Our findings - Correlation		○	○	○	○	○	○	
Belief on rel. performance	**Our findings - Full model**		∨	∨	∧	∧	∧	∧	∧
	Our findings - Correlation		∨	∨	○	∧	∧	∧	
	Buser et al. (2014)	✓					∧		
	Buser et al. (2022)¶	✓					∧		
	Gupta et al. (2011)	(✗)					○		
	Dohmen and Falk (2011)§	(✓)	∨		○			∧	
	Fornwagner et al. (2023)	✓					∧		
	Niederle and Vesterlund (2007)	✓					∧		
	Sutter and Glätzle-Rützler (2015)	✓					∧		
	Tungodden and Willén (2023)	✓					∧		
Overplacement	**Our findings - Full model**		∧	○	∨	∨	∨	∨	∨
	Our findings - Correlation		∧	○	∨	∨	∨	∨	
	Almås et al. (2016)	✗					∧		
	Reuben et al. (2017)	(✗)					○		
Patience	**Our findings - Full model**		∨	○	∧	○	○	○	○
	Our findings - Correlation		∨	○	○	○	○	∧	
	Almås et al. (2016)	○					∧		
Risk taking	**Our findings - Full model**		∨	○	○	∧	∧	∧	∧
	Our findings - Correlation		∨	∨	○	∧	∧	∧	
	Almås et al. (2016)	✓					∧		
	Buser et al. (2014)	✓					∧		

To be continued on next page ...

Variables	Our Findings/Literature	Consistent?	F ≺ V or T	V ≺ F or T	V ≺ F	T ≺ V or F	T ≺ V	T ≺ F	Ordered logit
	Buser et al. (2024)¶	✓					∧		
	Bonin et al. (2007)‡	✓					∧		
	Gupta et al. (2011)	✓					∧		
	Dohmen and Falk (2011)§	✓	∨			∧		∧	
	Eriksson et al. (2009)	✓					∧		
	Fornwagner et al. (2023)	(✗)					○		
	Niederle and Vesterlund (2007)	✓					∧		
	Sutter and Glätzle-Rützler (2015)	✓					∧		
	Tungodden and Willén (2023)	✓					∧		

Note: This table compares the results from Table A6 with the results from previous literature regarding the role of explanatory skill variables, demographics, personality traits and economic preferences on sorting decisions between the variable (V), fixed (F) and tournament payments (T). ∧ - Significant increase in likelihood. ∨ - Significant decrease in likelihood. ○ - No significant results. † - IQ proxied by GPA. Our findings are presented for both, the resp. coefficient for the full model on sorting decisions controlling for all other predictors and the plain correlation between predictor and sorting decision. ¶ - Investigates sorting decisions into college major choices and future earnings. § - Compared sorting decisions between fixed, variable, revenue sharing and tournament payments. $ - Compared sorting decisions for a linear and a convex payout schedule that paid more per correct answer. ‡ - Investigates sorting decisions into occupations with low earnings risk.

3.4 Identifying the Treatment Effect of Having a Choice

As a final part of this section, we isolate the overall treatment effect on performance of having agency over the payment scheme, and start by contrasting individuals' performance across the *Exogenous* and *Endogenous* treatment conditions. Looking at the performance conditional on the payment scheme depicted in Figure 2, and recalling the roughly equal choice frequencies of Fixed and Tournament payment (see Figure 3), we note that the average performance across all payments must practically be the same in *Exogenous* and *Endogenous*. In fact, the overall average performance is only 0.97 units lower, and insignificantly so, in *Endogenous* compared to *Exogenous* (121.77 vs. 122.74; t-test: $p = 0.37$). This means that allowing for self-selection into payment schemes is, on average, not beneficial for overall productivity. Of course, Figure 2 reveals that the overall null-effect does not apply to each payment scheme.

Yet, the samples under any of the three payment schemes in the *Endogeneous* treatment condition are not comparable to those in the respective payment scheme in the *Exogeneous* treatment condition, as we have seen in the last subsection, and the difference in performance observed between the treatments in Figure 2 is a mixture of having agency over the payment scheme, and different sample characteristics resulting from self selection (starting

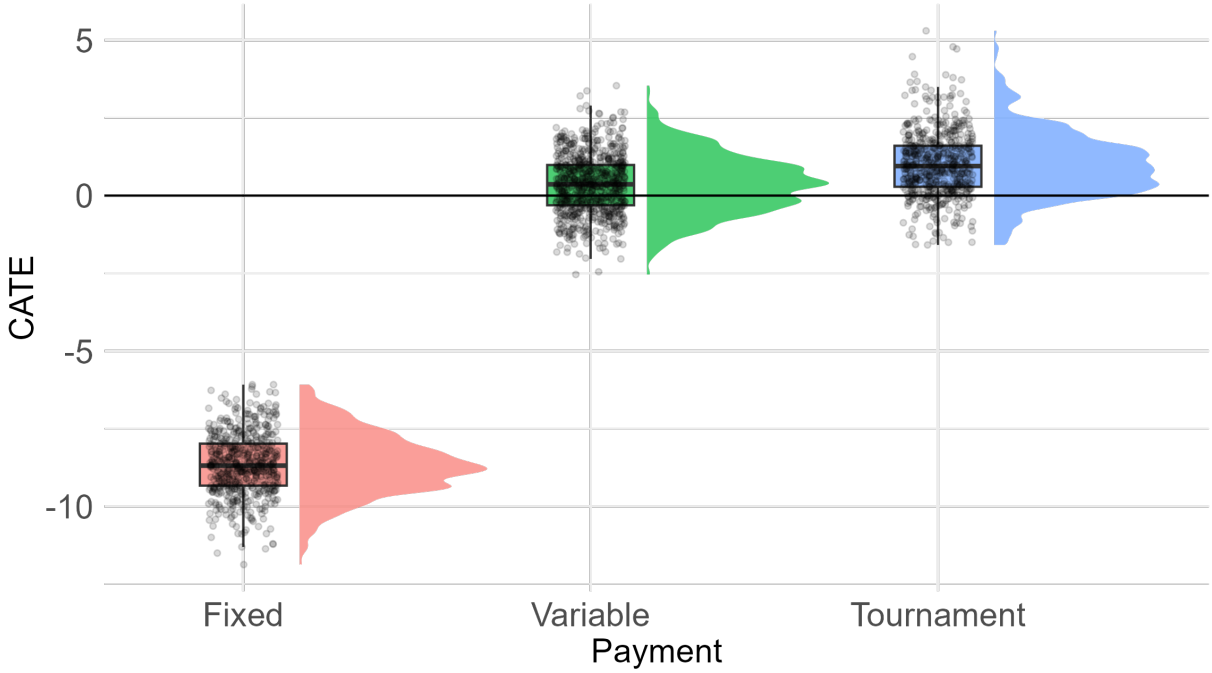
with productivity, as Figure 4 has shown).

To provide deeper insights, we use causal random forests (Wager and Athey, 2018). This type of machine learning algorithm takes into account the causal setting of experiments. While random forests predict outcomes on the level of fine-grid subgroups that are homogeneous in terms of baseline characteristics by taking the subgroup average, *causal* random forests go one step further. They predict outcomes for the scenario with and without having received the treatment, i.e., with and without having had the choice of a payment scheme for the cases where the treatment in fact was not received, and where it was received. Importantly, the most important individual characteristics and the payment scheme are kept constant in a given subgroup. Subgroups are defined in a data-driven way based on explanatory variables and corresponding cut-off values with the goal to maximize between-group heterogeneity of treatment effects (that is: groups are based to a higher degree on those variables that matter most for explaining the treatment effect such that subgroups are more homogeneous with respect to these variables). Aggregating the subgroup-based treatment effect estimations offers a way of obtaining an estimate for the conditional average treatment effects for the whole sample, or for certain subgroups, such as those that are linked to the different payment schemes considered here. Effectively, this allows us to compare similar participants in terms of individual characteristics who had self-selected themselves into a given payment scheme with those who had not, but instead have been randomly assigned – that is, we can keep everything (relevant) constant, except for having had the agency over a payment scheme or not, and in that way, "correct" for self-selection.

Figure 5 reports these conditional average treatment effects (CATE) conditional on the payment schemes. We find a negative CATE of having a choice regarding the payment scheme for the Fixed payment: individuals who had agency over the payment and chose the Fixed payment, had a profoundly lower performance than what they are predicted to have based on the performance of individuals with the same set of (important) characteristics who were randomly assigned the Fixed payment. For Variable and Tournament payment we find slightly positive but statistically insignificant differences between performance when having had a choice and performance when exogenously assigned to these payment schemes.

Hence, this exercise has shown that while most of the "performance gain" under Tournament payment in the *Endogenous* treatment that we have seen in Figure 2 is actually due to self-selection of more productive and more able individuals into this payment scheme (because, if we keep the samples comparable, as we do here, we see no significant differences anymore), the case under Fixed payment is different: There, only about half of the

Figure 5: Conditional Average Treatment Effects (CATEs) of Having Agency over the Choice of a Payment Scheme on Performance (by Payment Scheme).



Notes: This figure shows the distribution of CATEs within each payment scheme. Depicted boxplots show 25th/50th/75th percentiles and whiskers for 1.5×IQR (inter quartile range, the distance between 25th and 75th percentile). The CATE is obtained by comparing the actual with the counterfactual performance in the same payment scheme. Note that the counterfactual case is a prediction based on the causal random forest methodology (Wager and Athey, 2018) that predicts individuals' performance if the individual had been in the other treatment group and, thus, had (in *Exogenous*) or had not (in *Endogenous*) agency over their payment scheme. A CATE of zero indicates no treatment effect of having a choice regarding payment scheme on one's performance. A negative CATE indicates a negative treatment effect of having agency over payment schemes on performance.

"performance loss" (i.e., about 8 of 17 tasks solved less in *Endogenous*) can be attributed to different characteristics, while the other half can be attributed to being able to chose a payment scheme (Fixed payment) that pays the same amount irrespective of ability, effort, and luck. In fact, this negative effect of having a choice is reminiscent of similar findings in Adjerid et al. (2022) and Woerner et al. (forthcoming).

4 The Potential of Algorithmic Assignment to Payment Schemes

We now investigate whether and to which extent the own (endogenous) assignment to payment schemes could be improved, given the knowledge about which factors (with respect to personal traits, preferences, and skills) determine performance. Recall that Opitz et al. (2024) first ran an exogenous treatment – similar to ours – and then used the insights from this treatment for another exogenous assignment of a new set of MTurkers to the payment scheme that was predicted to yield the best output. In fact, they found that algorithmic

assignment improved performance beyond the single best payment scheme from their first treatment. Our approach is different, but complementary. We use machine learning algorithms to estimate whether we could have improved performance, earnings, and utility through such an assignment compared to participants' *own choices*. So, we do not run additional sessions with new (exogenous) treatments, but focus on our participants' own choices. Moreover, we consider three different outcomes, and in particular include outcomes that focus on the employee's perspective, too, which might be an important aspect to consider in light of potential backlash to the use of algorithmic tools in human resource practice (e.g., Park et al., 2021; Gonzalez et al., 2022).

For a measure of utility, we assume a standard utility function being defined as the difference between payoff and effort costs, $U(e) = \pi(e) - c(e)$ (see, e.g., DellaVigna and Pope, 2017; DellaVigna et al., 2022). To elicit the costs that correspond to the payoff for a given level of effort e in part 2, $c(e)$, we have asked all individuals how much effort they exerted, how stressed they felt, and how exhausted they got (all 1-7 Likert scales). These questions were asked right after the 20-minute real effort task in part 2. We define effort costs $c(e)$ as the mean of all three responses on the individual level standardized by the sample standard deviation across all three responses.

We run random forests trained on individuals in the *Exogenous* treatment to identify – without any sorting – influential predictors of our outcomes, i.e., performance, payoff, and utility, in each payment scheme. In a second step, we use these insights to predict each outcome for each individual in the *Endogenous* condition under the two unobserved payment schemes. This serves to get a prediction of each unobserved outcome if participants in *Endogenous* had been assigned exogenously to another payment scheme. Lastly, by comparing realized outcomes in the actually chosen payment scheme with our random forest predictions for a given set of characteristics we isolate one "optimal" payment scheme where each outcome is maximized on the individual level. The resulting differences between the realized outcomes in the actually chosen and the predicted outcome in the optimal payment condition – if the two do not coincide – allow us to quantify the discrepancies in outcomes due to possibly suboptimal sorting decisions. If the chosen and optimal payment schemes are identical, the difference is zero.

In the left panel of Figure 6, we show the results for performance, in the middle one for payoffs, and in the right one for utility. Within each panel, the first column refers to participants that have chosen the Fixed payment, the second to those in the Variable payment, and the third column to participants having chosen Tournament payment. In each column, we report which payment scheme would have been the optimal choice (for each of the three outcomes). In the first panel and the first column, we see, for example, that

for only 20% of participants who have chosen the Fixed payment, this was estimated to be the optimal choice for maximizing performance. For 48%, choosing the Variable payment would have led to higher performance, and about one third (32%) should have chosen the Tournament payment to maximize performance. For the group having chosen the Variable payment (second column), we note that 58% have maximized the estimated performance with this payment scheme (which is much better than the 20% for participants in the Fixed payment). For the group who chose the Tournament payment, almost two thirds (65%) are estimated to have taken the optimal choice.

The middle panel shows that with regards to potential earnings, the fraction of optimal choices is smaller than for performance in the first panel. This is due to the fact that the payment scheme with the highest estimated performance of an individual need not be the one with the highest earnings, because the latter depend on the absolute level of performance (for the Variable payment) and the relative performance compared to other participants (in the Tournament payment). Here, we see the largest fraction of optimal choice for the Tournament payment, where 52% of participants are estimated to have actually chosen the payment that maximizes their earnings.[20]

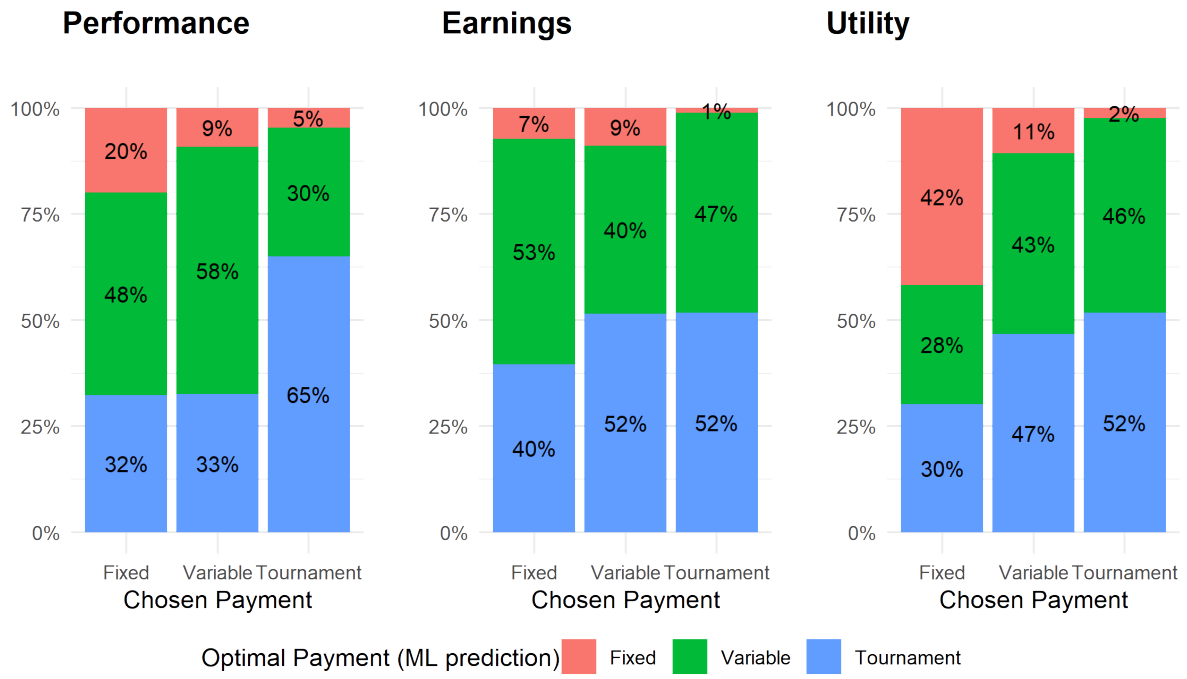
The right panel in Figure 6 shows the estimates for utility. Notably, 42% of participants who chose the Fixed payment are estimated to have optimized their utility (while performance was optimal only for 20% and earnings only for 7% of these participants). This shows that participants in the Fixed payment seem to have perceived the task as relatively costly, thus improving their utility by reducing effort. For subjects who have chosen the Variable or the Tournament payment, we see that they optimized their utility in 43%, respectively 52%, of cases.

Across all three outcome measures and all groups of participants (those choosing Fixed, Variable and Tournament payment), we observe that on average about 58.11% of participants make sorting decisions that fall short of their predicted potential.[21] This is a substantial share of participants, indicating a large potential for improving outcomes by taking into account personal characteristics when assigning subjects to different payment schemes. We

[20]Note that this prediction does not account for the arguably random "assignment" to low and high payment in the Tournament treatment, but predicts earnings solely based on skills, demographics, traits, preferences and the payment scheme. The given numbers are thus predictions for the earnings potentials on the individual level, where we did not implement additional assumptions to end up with a prediction on the realized earnings. The advantage of this approach is that, with even less assumptions, it already takes into account the probability of winning the tournament, given an individual's characteristics.

[21]58.11% results from adding all shares of participants that did not choose the outcome-maximizing payment scheme across all outcome-payment scheme combinations and dividing it by nine. We opt for the average, as it gives a good sense of the potential, although other aggregations might be adequate for different questions, such as, for example, which of the three dimension should ideally be targeted to implement an overall ideal choice.

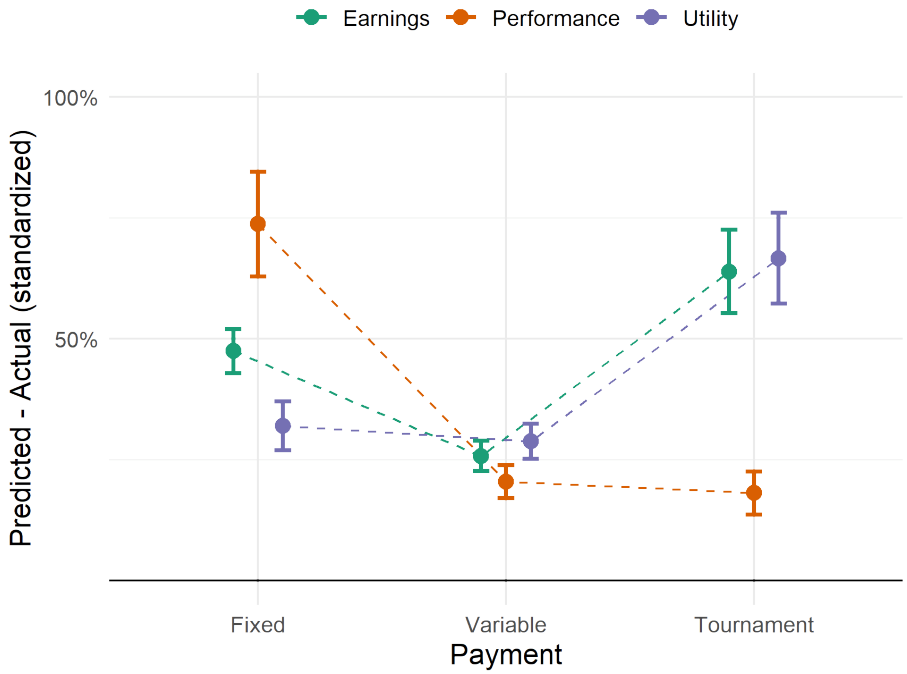
Figure 6: Algorithmic Assignment into Performance-, Earnings-, and Utility-Maximizing Payment Based on Random Forest Predictions Compared to Outcomes Resulting from Actual Choice



can quantify this potential as follows: Figure 7 reports results on the differences between actual and predicted outcomes. Note that we present differences in a standardized form, i.e., we divide differences by the sample standard deviation of the actual choice. Positive values indicate a potential for improvement. We find that the algorithmic assignment of payments is in each of the outcome dimensions, and for every payment scheme, able to assign individuals into payment schemes where the average predicted outcome is significantly higher than the actually realized average outcome.

While the previous analyses have shown that the potential for improvement is huge, in real life, it is of course only possible to optimize one dimension (or a linear combination of those, for example). In practice, we are thus interested whether we can improve even all outcomes on average with such an approach. Figure 8 shows that this is possible. The graph plots the differences between the actual outcome and the predicted outcome, when the utility-maximizing payment scheme is chosen (as opposed to above where, for every outcome, the prediction for the payment scheme that maximizes the given outcome has been selected). It is of course possible to select a more nuanced scoring function for the algorithmic assignment than our utility function (e.g., some weighted average of performance and utility), but since our utility function combines the trade-off between effort provision and performance via earnings, investigating the potential when selecting the utility-maximizing payment scheme serves as a good indicator for what is possible.

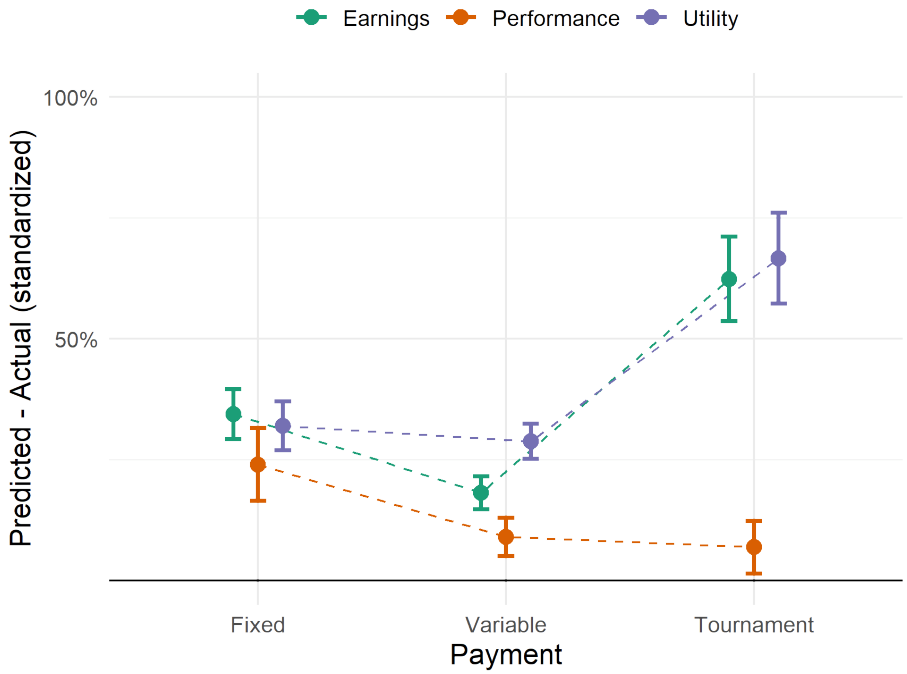
Figure 7: Potential Improvement through Algorithmic Assignment: Standardized Difference between Predicted and Actual Outcomes (Earnings, Performance, and Utility) across Payment Schemes.



Notes: Predicted values are based on algorithmic assignment into the outcome-maximizing payment scheme. Standardization results from dividing the difference between predicted and actual outcomes by the population standard deviation of the actual outcome. Brackets represent 95%-confidence intervals.

In Figure 8, we see in the third column that even for those who selected the Tournament payment, performance could have been improved significantly on average by assigning the utility-maximizing payment scheme. Earnings and Utility for this group could have been improved significantly by more than 50% of a standard deviation in the respective outcome. Hence, optimizing any scoring function that also targets performance directly (as opposed to indirectly, as is the case in our utility function), will even result in a larger improvement in performance for this group, while the improvement in Earnings and Utility will still be sizeable. For the other groups, i.e., those who selected the Fixed and the Variable payment (1st and 2nd column of Figure 8), we observe significant improvements between about 20 and 35% (Fixed payment) and about 10 to 30% of an SD in the respective outcome. Hence, such an approach would not only be beneficial for potential employers (who clearly prefer better performance), but also for the participants (with higher average earnings and utility).

Figure 8: Potential Improvement through Algorithmic Assignment When Optimization Targets Utility: Standardized Difference between Predicted and Actual Outcomes (Earnings, Performance, and Utility) across Payment Schemes.



Notes: Predicted values are based on algorithmic assignment into the payment scheme that would have optimized utility. Standardization results from dividing the difference between predicted and actual outcomes by the population standard deviation of the actual outcome. Brackets represent 95%-confidence intervals.

5 Conclusion

Improving performance through proper incentives sounds like a simple solution to the important question of how to motivate humans to perform at their best. Yet, it is not the case that one size fits all, nor does one payment scheme yield the best performance. Rather, humans react in very different ways to incentives, and this reaction depends on many factors, such as their abilities, background characteristics, personality traits, economic preferences and also their beliefs. For this reason, it is important to understand how these factors interact with different payment schemes for performance, earnings and utility from a task.

In this paper, we have studied in a lab-in-the-field experiment with 1,933 German high school students how a large plethora of factors influence sorting decisions and performance into Fixed, Variable and Tournament payment schemes. We have found evidence of systematic sorting across the three payment schemes based on socio-demographics, personality traits, preferences and beliefs. Interestingly, the factors that explain sorting are only partly overlapping with the factors that have explanatory power for performance. Moreover, the importance of the different factors also depends on which of the payment schemes are

compared to each other.

We have also shown that specific factors determine effort provision, but differently depending on the payment scheme. So, there is a large degree of heterogeneity in the reaction to incentives, which is the reason why one size cannot fit all. When given the choice to self-select into a specific payment scheme, we observe that individuals' choice behavior is often not in line with the factors that would maximize their effort provision, earnings or utility from the task. In fact, about 50% of sorting decisions could, on average, have been improved by a machine learning algorithm that is trained either on the relationship between personal characteristics and performance, or that between characteristics and earnings, or with utility when incentives are exogenously assigned. As abilities, beliefs about own abilities, socio-demographics, personality traits, and preferences are impacting sorting decisions and effort provision in different ways, depending on the payment scheme, it looks promising for future research to investigate these intricate interdependencies further. This may then help in raising awareness on how to improve labor market contracts, where the employer as well as the worker, can better account for the strengths and weaknesses of characteristics to optimize output and earnings, but also utility, under different payments schemes.

References

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

Adjerid, I., G. Loewenstein, R. Purta, and A. Striegel (2022). Gain-loss incentives and physical activity: The role of choice and wearable health tools. *Management Science 68*(4), 2642–2667.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Almås, I., A. W. Cappelen, K. G. Salvanes, E. O. Sørensen, and B. Tungodden (2016). Willingness to compete: Family matters. *Management Science 62*(8), 2149–2162.

Boneva, T., T. Buser, A. Falk, and F. Kosse (2022). The origins of gender differences in competitiveness and earnings expectations: Causal evidence from a mentoring intervention. *CEPR discussion paper no. DP17008*.

Bonin, H., T. Dohmen, A. Falk, D. Huffman, and U. Sunde (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics 14*(6), 926–937.

Borghans, L., H. Meijers, and B. Ter Weel (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry 46*(1), 2–12.

Buser, T., A. W. Cappelen, and B. Tungodden (2021). Fairness and willingness to compete. *NHH Dept. of Economics Discussion Paper* (08).

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., M. Niederle, and H. Oosterbeek (2024). Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. *Review of Economics and Statistics*, 1–45.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review 107*(5), 125–130.

Buser, T., N. Peter, and S. C. Wolter (2022). Willingness to compete, gender and career choices along the whole ability distribution. *Experimental Economics 25*(5), 1299–1326.

Cadena, B. C. and B. J. Keys (2015). Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy 7*(3), 126–153.

Cassar, L. and S. Meier (2018). Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives 32*(3), 215–38.

Castagnetti, A. and R. Schmacker (2022). Protecting the ego: Motivated information selection and updating. *European Economic Review 142*, 104007.

Charness, G., U. Gneezy, and A. Henderson (2018). Experimental methods: Measuring effort in Economics experiments. *Journal of Economic Behavior & Organization 149*, 74–87.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree – an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

DellaVigna, S., J. A. List, U. Malmendier, and G. Rao (2022). Estimating social preferences and gift exchange at work. *American Economic Review 112*(3), 1038–1074.

DellaVigna, S. and D. Pope (2017). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies 85*(2), 1029–1069.

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

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization 65*(3-4), 636–653.

Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky, and M. Vera-Hernández (2017). Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in India. *American Economic Review 107*(5), 506–510.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.

Duckworth, A. L. and P. D. Quinn (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment 91*(2), 166–174.

Eriksson, T., S. Teyssier, and M.-C. Villeval (2009). Self-selection and the efficiency of tournaments. *Economic Inquiry 47*(3), 530–548.

Essau, C. A., S. Sasagawa, and P. J. Frick (2006). Psychometric properties of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies 15*(5), 595–614.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science 69*(4), 1935–1950.

Fornwagner, H., M. Pompeo, and N. Serdarevic (2023). Choosing competition on behalf of someone else. *Management Science 69*(3), 1555–1574.

Frick, P. J. (1991). The Alabama Parenting Questionnaire. *University of Alabama*.

Gneezy, U., J. A. List, J. A. Livingston, X. Qin, S. Sadoff, and Y. Xu (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights 1*(3), 291–308.

Gneezy, U., S. Meier, and P. Rey-Biel (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives 25*(4), 191–210.

Gneezy, U. and P. Rey-Biel (2014). On the relative efficiency of performance pay and noncontingent incentives. *Journal of the European Economic Association 12*(1), 62–72.

Gonzalez, M. F., W. Liu, L. Shirase, D. L. Tomczak, C. E. Lobbe, R. Justenhoven, and N. R. Martin (2022). Allying with AI? Reactions toward human-based, AI/ML-based, and augmented hiring processes. *Computers in Human Behavior 130*, 107179.

Gupta, N. D., A. Poulsen, and M. C. Villeval (2011). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry 51*(1), 816–835.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences 104*(33), 13250–13255.

John, O. P. and S. Srivastava (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research*, Volume 2, pp. 102–138. Guilford.

Kosse, F. and M. M. Tincani (2020). Prosociality predicts labor market success around the world. *Nature Communications 11*(1), 5298.

Newby, J. L. and R. G. Klein (2014). Competitiveness reconceptualized: Psychometric development of the competitiveness orientation measure as a unified measure of trait competitiveness. *The Psychological Record 64*(4), 879–895.

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

Opitz, S., D. Sliwka, T. Vogelsang, and T. Zimmermann (2024). The algorithmic assignment of incentive schemes. *Management Science*, 1–18.

Park, H., D. Ahn, K. Hosanagar, and J. Lee (2021). Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. ACM.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature 37*(1), 7–63.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology 41*(1), 1–48.

Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.

Riener, G., S. O. Schneider, and V. Wagner (2020). Addressing validity and generalizability concerns in field experiments. *MPI Collective Goods Discussion Paper* (2020/16).

Schneider, S. O. and M. Schlather (2017). A new approach to treatment assignment for one and multiple treatment groups. *CRC Discussion Papers* (228).

Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science 58*(8), 1438–1457.

Sutter, M. and D. Glätzle-Rützler (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science 61*(10), 2339–2354.

Tungodden, J. and A. Willén (2023). When parents decide: Gender differences in competitiveness. *Journal of Political Economy 131*(3), 751–801.

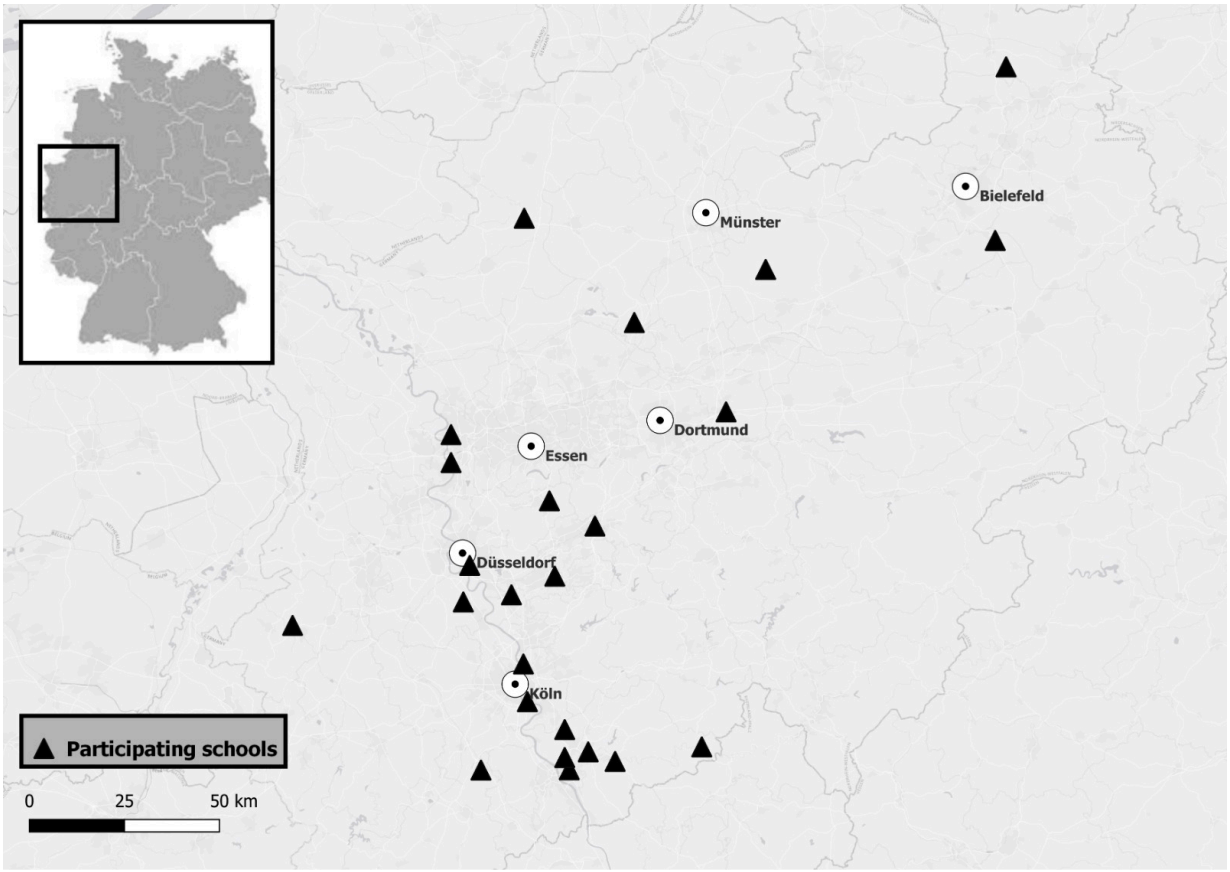
Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Woerner, A., G. Romagnoli, B. M. Probst, N. Bartmann, J. N. Cloughesy, and J. W. Lindemans (forthcoming). Should individuals choose their own incentives? Evidence from a mindfulness meditation intervention. *Management Science*.

A Appendix

A.1 Map of Participating Schools

Figure A1: Map of Participating Schools



A.2 Additional Results

Figure A2: Histogram of Solved Tables in Part 1 and 2.

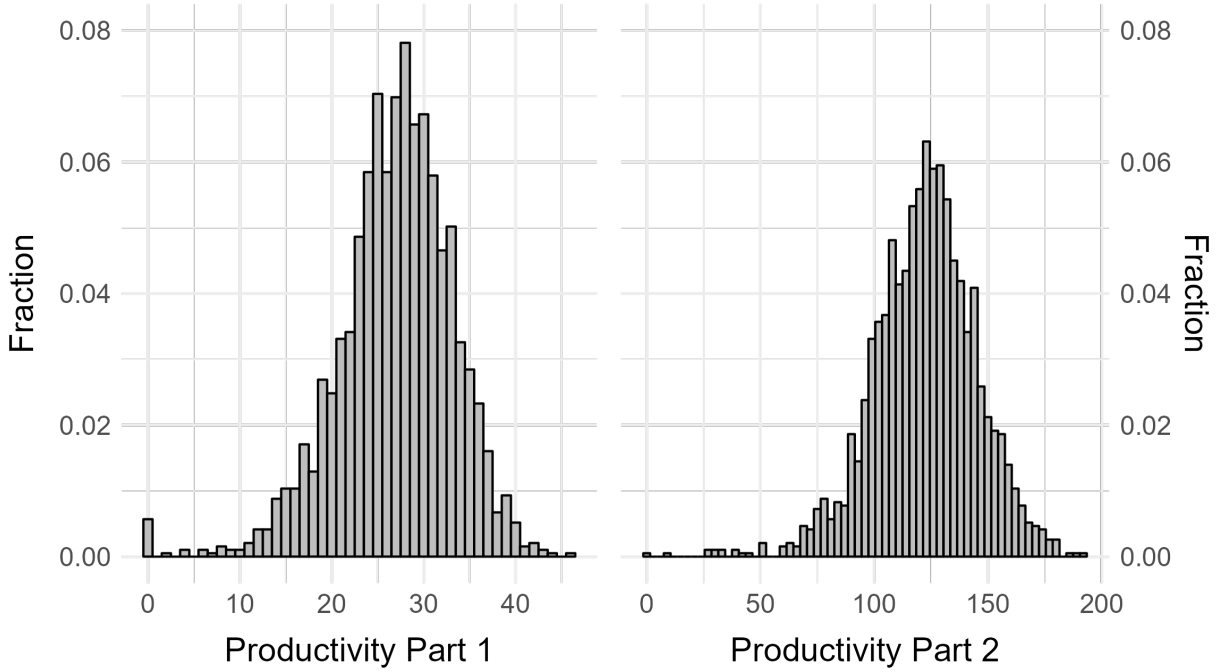


Table A1: Productivity in Exogenous (Reference Payment Scheme: Fixed): Interaction Effects Between Payment Schemes and Characteristics on Performance (Part 2)

	Performance
	(1)
Variable	15.874 (36.863)
Tournament	13.944 (25.599)
Productivity (resid.)	1.573 (0.271)***
Female (=1)	4.143 (2.630)
Age rel. to grade mean	-2.189 (0.761)***
Grade (9-13)	4.112 (0.795)***
IQ (Raven 0-10)	0.162 (0.525)
Grade Math	4.274 (0.894)***
Grade German	-1.113 (0.928)
Number of siblings	-0.260 (1.305)
Composite SES Index	-0.122 (0.867)
Patience Index	0.953 (1.104)
Risk Index	1.546 (1.141)
Altruism (0-10)	-0.392 (0.424)
Extraversion (1-5)	-1.220 (1.312)
Agreeableness (1-5)	1.662 (2.851)
Conscientiousness (1-5)	0.037 (2.318)
Neuroticism (1-5)	-0.250 (1.281)
Openness (1-5)	0.940 (1.528)
Enjoy Competition (1-5)	-0.392 (1.318)
Positive Parenting (1-5)	1.476 (1.356)
Grit (1-5)	-1.821 (1.866)
Belief on rel. performance (0-1)	30.142 (4.482)***
Overplacement (=1)	-14.325 (2.172)***
Variable x Productivity (resid.)	-0.173 (0.430)
Variable x Female (=1)	-2.263 (4.020)
Variable x Rel. age grade mean	-0.953 (1.585)
Variable x Grade (9-13)	-0.507 (1.884)
Variable x IQ (Raven 0-10)	0.194 (0.754)
Variable x Grade Math	-2.204 (0.993)**
Variable x Grade German	0.355 (1.962)
Variable x Number of siblings	-0.783 (1.763)
Variable x Composite SES Index	-0.504 (1.219)
Variable x Patience Index	-0.596 (1.523)
Variable x Risk Index	-2.660 (1.665)
Variable x Altruism (0-10)	1.029 (0.538)*
Variable x Extraversion (1-5)	1.596 (1.747)
Variable x Agreeableness (1-5)	-2.720 (3.511)
Variable x Conscientiousness (1-5)	0.776 (4.970)
Variable x Neuroticism (1-5)	-0.237 (1.923)
Variable x Openness (1-5)	-1.710 (1.935)
Variable x Enjoy Competition (1-5)	1.685 (2.103)
Variable x Positive Parenting (1-5)	-1.778 (1.949)
Variable x Grit (1-5)	1.183 (3.152)
Variable x Belief on rel. performance (0 - 1)	0.780 (5.311)
Variable x Overplacement (=1)	0.793 (2.190)
Tournament x Productivity (resid.)	0.140 (0.337)
Tournament x Female (=1)	-3.337 (3.256)
Tournament x Rel. age grade mean	0.674 (1.637)
Tournament x Grade (9-13)	0.118 (1.367)
Tournament x IQ (Raven 0-10)	0.782 (0.819)
Tournament x Grade Math	-3.161 (1.377)**
Tournament x Grade German	1.050 (1.451)
Tournament x Number of siblings	1.232 (1.490)
Tournament x Composite SES Index	-0.761 (1.428)
Tournament x Patience Index	-1.611 (1.812)
Tournament x Risk Index	-3.682 (1.801)**
Tournament x Altruism (0-10)	0.320 (0.686)
Tournament x Extraversion (1-5)	0.965 (2.276)
Tournament x Agreeableness (1-5)	-3.266 (3.055)
Tournament x Conscientiousness (1-5)	1.770 (3.104)
Tournament x Neuroticism (1-5)	-0.833 (2.215)
Tournament x Openness (1-5)	-0.534 (2.121)
Tournament x Enjoy Competition (1-5)	1.838 (1.828)
Tournament x Positive Parenting (1-5)	-3.092 (1.778)*
Tournament x Grit (1-5)	0.453 (2.901)
Tournament x Belief on rel. performance (0-1)	0.050 (4.968)
Tournament x Overplacement (=1)	-1.348 (2.242)
Constant	50.961 (20.672)**
Num.Obs.	983
R2 Adj.	0.347

Note: The table shows OLS regressions of performance (Part 2) on characteristics interacted with payment schemes. Standard errors (in parentheses) clustered on the session level.

Table A2: Productivity in Exogenous (Reference Payment Scheme: Variable): Interaction Effects Between Payment Schemes and Characteristics on Performance (Part 2)

	Performance
	(1)
Fixed	-15.874 (36.863)
Tournament	-1.930 (32.739)
Productivity (resid.)	1.400 (0.354)***
Female (=1)	1.881 (2.571)
Age rel. to grade mean	-3.142 (1.288)**
Grade (9-13)	3.605 (1.557)**
IQ (Raven 0-10)	0.356 (0.633)
Grade Math	2.070 (0.856)**
Grade German	-0.758 (1.781)
Number of siblings	-1.043 (0.828)
Composite SES Index	-0.626 (1.025)
Patience Index	0.357 (1.205)
Risk Index	-1.114 (1.158)
Altruism (0-10)	0.637 (0.477)
Extraversion (1-5)	0.376 (1.243)
Agreeableness (1-5)	-1.058 (1.975)
Conscientiousness (1-5)	0.813 (4.117)
Neuroticism (1-5)	-0.487 (1.322)
Openness (1-5)	-0.770 (1.916)
Enjoy Competition (1-5)	1.293 (1.258)
Positive Parenting (1-5)	-0.302 (1.065)
Grit (1-5)	-0.638 (2.415)
Belief on rel. performance (0-1)	30.923 (4.022)***
Overplacement (=1)	-13.532 (2.428)***
Fixed x Productivity (resid.)	0.173 (0.430)
Fixed x Female (=1)	2.263 (4.020)
Fixed x Rel. age grade mean	0.953 (1.585)
Fixed x Grade (9-13)	0.507 (1.884)
Fixed x IQ (Raven 0-10)	-0.194 (0.754)
Fixed x Grade Math	2.204 (0.993)**
Fixed x Grade German	-0.355 (1.962)
Fixed x Number of siblings	0.783 (1.763)
Fixed x Composite SES Index	0.504 (1.219)
Fixed x Patience Index	0.596 (1.523)
Fixed x Risk Index	2.660 (1.665)
Fixed x Altruism (0-10)	-1.029 (0.538)*
Fixed x Extraversion (1-5)	-1.596 (1.747)
Fixed x Agreeableness (1-5)	2.720 (3.511)
Fixed x Conscientiousness (1-5)	-0.776 (4.970)
Fixed x Neuroticism (1-5)	0.237 (1.923)
Fixed x Openness (1-5)	1.710 (1.935)
Fixed x Enjoy Competition (1-5)	-1.685 (2.103)
Fixed x Positive Parenting (1-5)	1.778 (1.949)
Fixed x Grit (1-5)	-1.183 (3.152)
Fixed x Belief on rel. performance (0-1)	-0.780 (5.311)
Fixed x Overplacement (=1)	-0.793 (2.190)
Tournament x Productivity (resid.)	0.313 (0.505)
Tournament x Female (=1)	-1.075 (4.186)
Tournament x Rel. age grade mean	1.626 (1.872)
Tournament x Grade (9-13)	0.625 (1.350)
Tournament x IQ (Raven 0-10)	0.588 (0.773)
Tournament x Grade Math	-0.957 (1.342)
Tournament x Grade German	0.696 (2.234)
Tournament x Number of siblings	2.015 (1.149)*
Tournament x Composite SES Index	-0.257 (1.594)
Tournament x Patience Index	-1.015 (1.736)
Tournament x Risk Index	-1.022 (2.009)
Tournament x Altruism (0-10)	-0.709 (0.661)
Tournament x Extraversion (1-5)	-0.630 (2.127)
Tournament x Agreeableness (1-5)	-0.546 (2.964)
Tournament x Conscientiousness (1-5)	0.995 (5.008)
Tournament x Neuroticism (1-5)	-0.596 (1.755)
Tournament x Openness (1-5)	1.176 (2.661)
Tournament x Enjoy Competition (1-5)	0.153 (1.781)
Tournament x Positive Parenting (1-5)	-1.314 (1.482)
Tournament x Grit (1-5)	-0.730 (3.354)
Tournament x Belief on rel. performance (0-1)	-0.730 (5.926)
Tournament x Overplacement (=1)	-2.141 (2.313)
Constant	66.835 (26.020)**
Num.Obs.	983
R2 Adj.	0.347

Note: The table shows OLS regressions of performance (Part 2) on characteristics interacted with payment schemes. Standard errors (in parentheses) clustered on the session level.

Table A3: Productivity in Endogenous (Reference Payment Scheme: Fixed): Interaction Effects Between Payments and Characteristics on Performance (Part 2)

	Performance
	(1)
Variable	-22.923 (29.687)
Tournament	-20.208 (26.158)
Productivity (resid.)	0.786 (0.751)
Female (=1)	0.247 (5.498)
Age rel. to grade mean	1.420 (1.433)
Grade (9-13)	1.084 (1.930)
IQ (Raven 0-10)	1.132 (1.384)
Grade Math	2.018 (1.255)
Grade German	-4.443 (1.537)***
Number of siblings	-1.540 (1.058)
Composite SES Index	0.541 (1.373)
Patience Index	-2.908 (2.196)
Risk Index	-0.243 (2.691)
Altruism (0-10)	0.845 (1.038)
Extraversion (1-5)	-4.319 (2.284)*
Agreeableness (1-5)	-3.319 (4.093)
Conscientiousness (1-5)	6.459 (5.429)
Neuroticism (1-5)	-1.422 (3.360)
Openness (1-5)	0.485 (2.629)
Enjoy Competition (1-5)	-1.160 (1.430)
Positive Parenting (1-5)	-1.816 (2.105)
Grit (1-5)	-1.539 (4.779)
Belief on rel. performance (0-1)	34.619 (7.543)***
Overplacement (=1)	-16.678 (3.006)***
Variable x Productivity (resid.)	1.466 (0.714)**
Variable x Female (=1)	3.123 (5.722)
Variable x Rel. age grade mean	-1.338 (1.494)
Variable x Grade (9-13)	0.581 (2.210)
Variable x IQ (Raven 0-10)	-0.301 (1.288)
Variable x Grade Math	-0.280 (1.501)
Variable x Grade German	3.081 (1.825)*
Variable x Number of siblings	1.607 (1.018)
Variable x Composite SES Index	-1.354 (1.527)
Variable x Patience Index	2.764 (2.273)
Variable x Risk Index	0.526 (2.831)
Variable x Altruism (0-10)	-0.696 (1.220)
Variable x Extraversion (1-5)	4.851 (2.147)**
Variable x Agreeableness (1-5)	6.461 (4.656)
Variable x Conscientiousness (1-5)	-3.544 (6.120)
Variable x Neuroticism (1-5)	2.120 (3.496)
Variable x Openness (1-5)	-2.852 (3.098)
Variable x Enjoy Competition (1-5)	1.891 (1.260)
Variable x Positive Parenting (1-5)	1.562 (1.942)
Variable x Grit (1-5)	-1.830 (5.332)
Variable x Belief on rel. performance (0 - 1)	-4.665 (8.230)
Variable x Overplacement (=1)	-4.063 (3.018)
Tournament x Productivity (resid.)	1.529 (0.573)***
Tournament x Female (=1)	2.146 (6.208)
Tournament x Rel. age grade mean	-1.477 (1.573)
Tournament x Grade (9-13)	1.173 (2.300)
Tournament x IQ (Raven 0-10)	0.195 (1.990)
Tournament x Grade Math	-1.179 (1.563)
Tournament x Grade German	5.478 (2.108)***
Tournament x Number of siblings	2.210 (1.689)
Tournament x Composite SES Index	1.132 (1.608)
Tournament x Patience Index	3.337 (3.027)
Tournament x Risk Index	-2.185 (2.711)
Tournament x Altruism (0-10)	-1.506 (0.905)*
Tournament x Extraversion (1-5)	6.498 (2.582)**
Tournament x Agreeableness (1-5)	5.924 (4.603)
Tournament x Conscientiousness (1-5)	-13.096 (6.369)**
Tournament x Neuroticism (1-5)	0.260 (3.244)
Tournament x Openness (1-5)	-1.175 (3.208)
Tournament x Enjoy Competition (1-5)	0.309 (2.173)
Tournament x Positive Parenting (1-5)	2.215 (2.164)
Tournament x Grit (1-5)	4.470 (5.575)
Tournament x Belief on rel. performance (0-1)	-7.541 (8.024)
Tournament x Overplacement (=1)	-0.129 (3.748)
Constant	108.566 (21.650)***
Num.Obs.	950
R2 Adj.	0.495

Note: The table shows OLS regressions of performance (Part 2) on characteristics interacted with payment schemes. Standard errors (in parentheses) clustered on the session level.

Table A4: Productivity in Endogenous (Reference Payment Scheme: Variable): Interaction Effects Between Payments and Characteristics on Performance (Part 2)

	Performance
	(1)
Fixed	22.923 (29.687)
Tournament	2.715 (24.411)
Productivity (resid.)	2.252 (0.319)***
Female (=1)	3.370 (1.559)**
Age rel. to grade mean	0.081 (0.983)
Grade (9-13)	1.665 (0.831)**
IQ (Raven 0-10)	0.830 (0.509)
Grade Math	1.738 (0.773)**
Grade German	-1.362 (0.960)
Number of siblings	0.067 (0.731)
Composite SES Index	-0.813 (0.496)
Patience Index	-0.144 (0.777)
Risk Index	0.283 (0.794)
Altruism (0-10)	0.149 (0.363)
Extraversion (1-5)	0.532 (1.406)
Agreeableness (1-5)	3.141 (1.555)**
Conscientiousness (1-5)	2.915 (2.318)
Neuroticism (1-5)	0.698 (1.063)
Openness (1-5)	-2.367 (1.346)*
Enjoy Competition (1-5)	0.731 (0.702)
Positive Parenting (1-5)	-0.254 (0.670)
Grit (1-5)	-3.369 (1.491)**
Belief on rel. performance (0-1)	29.954 (3.916)***
Overplacement (=1)	-20.741 (1.365)***
Fixed x Productivity (resid.)	-1.466 (0.714)**
Fixed x Female (=1)	-3.123 (5.722)
Fixed x Rel. age grade mean	1.338 (1.494)
Fixed x Grade (9-13)	-0.581 (2.210)
Fixed x IQ (Raven 0-10)	0.301 (1.288)
Fixed x Grade Math	0.280 (1.501)
Fixed x Grade German	-3.081 (1.825)*
Fixed x Number of siblings	-1.607 (1.018)
Fixed x Composite SES Index	1.354 (1.527)
Fixed x Patience Index	-2.764 (2.273)
Fixed x Risk Index	-0.526 (2.831)
Fixed x Altruism (0-10)	0.696 (1.220)
Fixed x Extraversion (1-5)	-4.851 (2.147)**
Fixed x Agreeableness (1-5)	-6.461 (4.656)
Fixed x Conscientiousness (1-5)	3.544 (6.120)
Fixed x Neuroticism (1-5)	-2.120 (3.496)
Fixed x Openness (1-5)	2.852 (3.098)
Fixed x Enjoy Competition (1-5)	-1.891 (1.260)
Fixed x Positive Parenting (1-5)	-1.562 (1.942)
Fixed x Grit (1-5)	1.830 (5.332)
Fixed x Belief on rel. performance (0-1)	4.665 (8.230)
Fixed x Overplacement (=1)	4.063 (3.018)
Tournament x Productivity (resid.)	0.063 (0.508)
Tournament x Female (=1)	-0.977 (4.206)
Tournament x Rel. age grade mean	-0.138 (1.353)
Tournament x Grade (9-13)	0.592 (1.483)
Tournament x IQ (Raven 0-10)	0.496 (1.046)
Tournament x Grade Math	-0.899 (1.450)
Tournament x Grade German	2.396 (1.687)
Tournament x Number of siblings	0.603 (1.421)
Tournament x Composite SES Index	2.486 (1.144)**
Tournament x Patience Index	0.573 (1.681)
Tournament x Risk Index	-2.711 (1.530)*
Tournament x Altruism (0-10)	-0.810 (0.600)
Tournament x Extraversion (1-5)	1.647 (1.873)
Tournament x Agreeableness (1-5)	-0.536 (2.207)
Tournament x Conscientiousness (1-5)	-9.552 (4.144)**
Tournament x Neuroticism (1-5)	-1.860 (2.106)
Tournament x Openness (1-5)	1.677 (1.877)
Tournament x Enjoy Competition (1-5)	-1.581 (1.641)
Tournament x Positive Parenting (1-5)	0.653 (1.231)
Tournament x Grit (1-5)	6.300 (2.782)**
Tournament x Belief on rel. performance (0-1)	-2.876 (4.601)
Tournament x Overplacement (=1)	3.934 (2.591)
Constant	85.643 (14.609)***
Num.Obs.	950
R2 Adj.	0.495

Note: The table shows OLS regressions of performance (Part 2) on characteristics interacted with payment schemes. Standard errors (in parentheses) clustered on the session level.

Table A5: Determinants of beliefs on rel. self-assessment and overplacement

	Belief on rel. performance (0-1)	**Overplacement (=1)**
	(I)	(II)
Skills		
Grade German	0.000	-0.170***
	(0.007)	(0.059)
Grade Math	0.009*	-0.014
	(0.005)	(0.052)
IQ (Raven 0-10)	-0.001	0.006
	(0.003)	(0.034)
Productivity (resid.)	0.000	0.000
	(0.002)	(0.016)
Demographics		
Age rel. to grade mean	0.007	0.146***
	(0.008)	(0.051)
Female (=1)	-0.027**	-0.246**
	(0.013)	(0.104)
Grade (9-13)	0.009	-0.019
	(0.006)	(0.060)
Number of siblings	0.001	0.002
	(0.005)	(0.043)
Positive Parenting (1-5)	0.011*	0.197***
	(0.006)	(0.061)
Composite SES Index	0.015***	0.051
	(0.006)	(0.046)
Personality Traits, Economic Preferences, and Beliefs		
Altruism (0-10)	0.001	-0.023
	(0.003)	(0.024)
Extraversion (1-5)	-0.004	0.112
	(0.007)	(0.072)
Agreeableness (1-5)	0.001	0.100
	(0.011)	(0.107)
Conscientiousness (1-5)	-0.003	-0.009
	(0.018)	(0.175)
Neuroticism (1-5)	-0.014	0.007
	(0.010)	(0.078)
Openness (1-5)	0.000	-0.135**
	(0.011)	(0.067)
Enjoy Competition (1-5)	0.000	-0.138***
	(0.006)	(0.051)
Grit (1-5)	0.002	0.050
	(0.016)	(0.134)
Patience Index	-0.002	-0.049
	(0.008)	(0.063)
Risk Taking Index	0.020***	0.029
	(0.007)	(0.066)
Constant	0.433***	0.954
	(0.104)	(1.045)
Num.Obs.	1933	1933
R2 Adj.	0.008	
RMSE	0.24	0.49

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

Note: Predictors of beliefs about own rel. performance in part 1 and overplacement (both normalized on the session level). Overplacement represents a dummy variable for a positive difference between the belief about own relative performance and actual own relative performance in part 1. Clustered standard errors on the session level

Table A6: Choice Regressions

	Logit						Ordered Logit
	Choice of Payment Scheme						
	F vs. V or T	V vs. F or T	V vs. F	T vs. V or F	T vs. V	T vs. F	F – V – T
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Skills							
Grade German	-0.193**	0.099	0.196**	0.070	0.018	0.137	0.142**
	(0.083)	(0.101)	(0.100)	(0.105)	(0.128)	(0.104)	(0.066)
Grade Math	-0.101	-0.099	0.035	0.252***	0.226**	0.308***	0.156**
	(0.073)	(0.060)	(0.078)	(0.093)	(0.097)	(0.110)	(0.066)
IQ (Raven 0-10)	-0.028	-0.021	-0.003	0.055	0.075	0.059	0.044
	(0.065)	(0.050)	(0.066)	(0.058)	(0.063)	(0.086)	(0.049)
Productivity (resid.)	-0.097***	0.018	0.093***	0.096***	0.064*	0.135**	0.102***
	(0.019)	(0.015)	(0.018)	(0.036)	(0.037)	(0.054)	(0.020)
Demographics							
Age rel. to grade mean	-0.039	0.059	0.055	-0.075	-0.101	-0.013	-0.008
	(0.146)	(0.104)	(0.148)	(0.112)	(0.129)	(0.182)	(0.101)
Female (=1)	-0.367**	0.682***	0.540***	-0.617**	-0.770***	-0.189	-0.128
	(0.179)	(0.151)	(0.168)	(0.259)	(0.259)	(0.322)	(0.201)
Grade (9-13)	0.127	-0.199*	-0.170	0.117	0.179*	-0.056	0.015
	(0.145)	(0.104)	(0.145)	(0.096)	(0.106)	(0.152)	(0.096)
Number of siblings	0.034	-0.029	-0.033	0.028	0.039	-0.013	-0.010
	(0.069)	(0.047)	(0.067)	(0.067)	(0.065)	(0.112)	(0.052)
Positive Parenting (0-5)	0.129	-0.041	-0.119	-0.083	-0.041	-0.180	-0.099
	(0.107)	(0.072)	(0.105)	(0.100)	(0.098)	(0.149)	(0.083)
Composite SES Index	-0.027	0.023	0.019	-0.023	-0.012	-0.068	0.015
	(0.096)	(0.068)	(0.094)	(0.073)	(0.084)	(0.117)	(0.072)
Personality Traits, Economic Preferences, and Beliefs							
Altruism (0-10)	-0.027	0.047	0.047	-0.018	-0.035	0.031	0.000
	(0.036)	(0.037)	(0.038)	(0.049)	(0.053)	(0.059)	(0.037)
Extraversion (0-5)	0.242**	-0.168*	-0.264**	0.015	0.055	-0.115	-0.133
	(0.118)	(0.101)	(0.126)	(0.103)	(0.112)	(0.154)	(0.082)
Agreeableness (0-5)	0.073	0.095	-0.001	-0.204	-0.197	-0.295	-0.117
	(0.153)	(0.142)	(0.161)	(0.176)	(0.181)	(0.214)	(0.130)
Conscientiousness (0-5)	-0.188	0.260	0.210	-0.208	-0.315	0.295	-0.044
	(0.254)	(0.233)	(0.264)	(0.244)	(0.263)	(0.313)	(0.185)
Neuroticism (0-5)	0.260**	-0.169*	-0.282**	-0.020	0.050	-0.144	-0.157
	(0.126)	(0.102)	(0.131)	(0.156)	(0.150)	(0.205)	(0.118)
Openness (0-5)	0.165	-0.087	-0.136	0.000	0.108	-0.223	-0.079
	(0.122)	(0.089)	(0.117)	(0.145)	(0.151)	(0.199)	(0.113)
Enjoy Competition (1-5)	-0.382***	-0.109	0.232***	0.577***	0.499***	0.762***	0.456***
	(0.072)	(0.087)	(0.083)	(0.093)	(0.104)	(0.117)	(0.061)
Grit (1-5)	0.164	-0.126	-0.203	0.041	0.157	-0.335	-0.027
	(0.213)	(0.175)	(0.228)	(0.211)	(0.224)	(0.250)	(0.185)
Belief on rel. performance (0-1)	-1.107***	-0.410**	0.693**	1.636***	1.376***	1.936***	1.379***
	(0.293)	(0.207)	(0.313)	(0.288)	(0.272)	(0.413)	(0.237)
Overplacement (=1)	0.799***	-0.052	-0.661***	-0.731***	-0.507***	-1.211***	-0.742***
	(0.191)	(0.129)	(0.200)	(0.196)	(0.197)	(0.248)	(0.150)
Patience Index	-0.172**	0.045	0.144*	0.102	0.021	0.180	0.132
	(0.074)	(0.086)	(0.079)	(0.137)	(0.143)	(0.142)	(0.089)
Risk Taking Index	-0.244**	-0.058	0.151	0.324**	0.307**	0.421**	0.298***
	(0.109)	(0.099)	(0.109)	(0.131)	(0.144)	(0.178)	(0.098)
Constant	-2.381	2.895*	3.019	-4.496***	-4.995***	-1.088	
	(2.103)	(1.685)	(2.262)	(1.084)	(1.455)	(1.830)	
Num.Obs.	950	950	693	950	715	492	950
RMSE	0.41	0.48	0.45	0.39	0.43	0.41	1.84

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

Note: Models (I) - (VI) are logit regressions. Model (I) compares choosing the Fixed (F) over the Variable (V) or Tournament (T) payment. Model (II) compares choosing the Variable over the Fixed or Tournament payment. Model (III) compares choosing the Variable over the Fixed payment. Model (IV) compares choosing the Tournament over the Fixed or Variable payment. Model (V) compares choosing the Tournament over the Variable payment. Model (VI) compares choosing the Tournament over the Fixed payment. Model (VII) is an ordered logit regression (F – V – T). Clustered standard errors on the session level in parentheses.

Table A7: Productivity by Treatment and Payment Scheme – Controlling for the Number of Days between Part 1 and Part 2 of the Experiment

	Exogenous			Endogenous		
	Payment Scheme					
	Fixed	Variable	Tournament	Fixed	Variable	Tournament
	(I)	(II)	(III)	(IV)	(V)	(VI)
Skills						
Grade German	-1.128 (0.924)	-0.793 (1.757)	-0.198 (1.189)	-4.436 (1.575)***	-1.361 (0.951)	1.017 (1.194)
Grade Math	4.142 (0.902)***	2.114 (0.870)**	1.246 (0.931)	2.127 (1.290)	1.738 (0.767)**	0.843 (1.102)
IQ (Raven 0-10)	0.156 (0.532)	0.371 (0.622)	0.884 (0.578)	1.160 (1.438)	0.831 (0.503)*	1.324 (0.880)
Productivity (resid.)	1.597 (0.273)***	1.407 (0.347)***	1.717 (0.404)***	0.795 (0.766)	2.252 (0.316)***	2.317 (0.385)***
Demographics						
Age rel. to grade mean	-2.363 (0.750)***	-3.141 (1.303)**	-1.461 (1.206)	0.822 (1.368)	0.080 (0.966)	-0.056 (1.140)
Female (=1)	4.046 (2.630)	1.877 (2.571)	0.610 (2.313)	0.127 (5.565)	3.369 (1.543)**	2.430 (3.386)
Grade (9-13)	4.008 (0.835)***	3.631 (1.549)**	4.441 (0.941)***	1.052 (1.887)	1.669 (0.814)**	2.251 (1.518)
Number of siblings	-0.226 (1.318)	-1.049 (0.822)	0.746 (0.767)	-1.725 (1.099)	0.067 (0.727)	0.678 (1.406)
Positive Parenting (0-5)	1.513 (1.356)	-0.325 (1.072)	-1.591 (1.095)	-2.085 (2.247)	-0.254 (0.664)	0.398 (1.105)
Composite SES Index	-0.081 (0.845)	-0.643 (1.049)	-0.893 (1.125)	0.650 (1.378)	-0.814 (0.493)*	1.678 (1.189)
Personality Traits, Economic Preferences, and Beliefs						
Altruism (0-10)	-0.376 (0.421)	0.657 (0.504)	0.003 (0.590)	0.950 (1.046)	0.149 (0.360)	-0.663 (0.411)
Extraversion (0-5)	-1.131 (1.316)	0.425 (1.320)	-0.379 (1.558)	-4.209 (2.268)*	0.531 (1.400)	2.185 (1.323)
Agreeableness (0-5)	1.734 (2.833)	-1.050 (1.976)	-1.257 (1.885)	-3.104 (4.212)	3.142 (1.544)**	2.611 (1.493)*
Conscientiousness (0-5)	-0.034 (2.367)	0.809 (4.121)	1.668 (2.812)	6.724 (5.434)	2.914 (2.283)	-6.629 (3.287)**
Neuroticism (0-5)	-0.324 (1.263)	-0.447 (1.304)	-0.844 (1.495)	-1.300 (3.372)	0.697 (1.059)	-1.167 (1.730)
Openness (0-5)	0.970 (1.547)	-0.781 (1.929)	0.421 (1.384)	0.387 (2.693)	-2.365 (1.337)*	-0.714 (1.386)
Enjoy Competition (1-5)	-0.409 (1.332)	1.272 (1.281)	1.494 (1.320)	-1.121 (1.489)	0.732 (0.701)	-0.845 (1.338)
Grit (1-5)	-1.634 (1.887)	-0.700 (2.465)	-1.344 (2.098)	-1.929 (4.761)	-3.369 (1.475)**	2.928 (2.205)
Belief on rel. performance (0-1)	30.265 (4.467)***	30.930 (4.046)***	30.248 (4.099)***	34.185 (7.785)***	29.957 (3.870)***	27.062 (3.535)***
Overplacement (=1)	-14.321 (2.157)***	-13.490 (2.485)***	-15.726 (1.813)***	-16.401 (3.015)***	-20.740 (1.352)***	-16.808 (2.110)***
Patience Index	0.939 (1.098)	0.349 (1.206)	-0.545 (1.503)	-3.022 (2.232)	-0.143 (0.768)	0.436 (1.245)
Risk Taking Index	1.388 (1.201)	-1.122 (1.172)	-1.839 (1.613)	-0.036 (2.765)	0.284 (0.789)	-2.428 (1.189)**
Time Difference between Sessions (in days)						
Time Difference	0.149 (0.103)	-0.052 (0.185)	-0.238 (0.139)*	-0.238 (0.221)	-0.002 (0.092)	0.013 (0.168)
Constant	49.690 (20.796)**	67.048 (25.786)***	64.615 (18.046)***	111.165 (20.467)***	85.612 (14.342)***	88.368 (23.537)***
Num.Obs.	331	327	325	235	458	257
R2 Adj.	0.343	0.341	0.345	0.135	0.476	0.470

Note: Table shows OLS regressions of performance in part 2 split by treatment and payment scheme. Standard errors (in parentheses) clustered on the session level.

Table A8: Choice Regressions – Controlling for the Number of Days between Part 1 and Part 2 of the Experiment

	Logit						Ordered Logit
	Choice of Payment Scheme						
	F vs. V or T	V vs. F or T	V vs. F	T vs. V or F	T vs. V	T vs. F	F – V – T
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Skills							
Grade German	-0.205**	0.105	0.198**	0.070	0.014	0.146	0.146**
	(0.086)	(0.100)	(0.099)	(0.105)	(0.128)	(0.105)	(0.066)
Grade Math	-0.100	-0.099	0.040	0.252***	0.227**	0.308***	0.157**
	(0.073)	(0.063)	(0.084)	(0.093)	(0.099)	(0.109)	(0.063)
IQ (Raven 0-10)	-0.035	-0.018	0.002	0.055	0.074	0.063	0.046
	(0.064)	(0.049)	(0.065)	(0.058)	(0.064)	(0.087)	(0.049)
Productivity (resid.)	-0.097***	0.018	0.096***	0.096***	0.064*	0.133**	0.103***
	(0.019)	(0.015)	(0.019)	(0.036)	(0.037)	(0.054)	(0.020)
Demographics							
Age rel. to grade mean	-0.012	0.036	0.009	-0.075	-0.098	-0.032	-0.025
	(0.158)	(0.101)	(0.158)	(0.113)	(0.124)	(0.194)	(0.114)
Female (=1)	-0.343*	0.672***	0.532***	-0.617**	-0.763***	-0.203	-0.137
	(0.186)	(0.155)	(0.172)	(0.259)	(0.261)	(0.328)	(0.202)
Grade (9-13)	0.101	-0.174*	-0.131	0.117	0.169	-0.054	0.039
	(0.141)	(0.096)	(0.139)	(0.099)	(0.109)	(0.149)	(0.101)
Number of siblings	0.042	-0.036	-0.043	0.028	0.040	-0.021	-0.015
	(0.069)	(0.047)	(0.066)	(0.069)	(0.066)	(0.112)	(0.053)
Positive Parenting (0-5)	0.136	-0.047	-0.133	-0.083	-0.041	-0.184	-0.103
	(0.106)	(0.073)	(0.105)	(0.099)	(0.098)	(0.149)	(0.083)
Composite SES Index	-0.023	0.019	0.017	-0.023	-0.008	-0.065	0.010
	(0.094)	(0.065)	(0.090)	(0.073)	(0.085)	(0.115)	(0.071)
Personality Traits, Economic Preferences, and Beliefs							
Altruism (0-10)	-0.033	0.050	0.052	-0.018	-0.035	0.034	0.002
	(0.036)	(0.038)	(0.039)	(0.050)	(0.053)	(0.061)	(0.037)
Extraversion (0-5)	0.254**	-0.173*	-0.273**	0.014	0.058	-0.121	-0.137*
	(0.121)	(0.101)	(0.127)	(0.103)	(0.113)	(0.155)	(0.082)
Agreeableness (0-5)	0.079	0.091	0.004	-0.204	-0.198	-0.294	-0.119
	(0.156)	(0.138)	(0.162)	(0.174)	(0.180)	(0.217)	(0.131)
Conscientiousness (0-5)	-0.184	0.252	0.198	-0.208	-0.309	0.301	-0.053
	(0.265)	(0.237)	(0.273)	(0.244)	(0.263)	(0.306)	(0.189)
Neuroticism (0-5)	0.264**	-0.170*	-0.291**	-0.020	0.050	-0.138	-0.160
	(0.125)	(0.102)	(0.129)	(0.156)	(0.151)	(0.202)	(0.119)
Openness (0-5)	0.148	-0.069	-0.127	0.000	0.099	-0.215	-0.065
	(0.122)	(0.091)	(0.118)	(0.140)	(0.145)	(0.197)	(0.109)
Enjoy Competition (1-5)	-0.386***	-0.109	0.244***	0.577***	0.498***	0.760***	0.460***
	(0.074)	(0.087)	(0.083)	(0.093)	(0.104)	(0.118)	(0.063)
Grit (1-5)	0.158	-0.128	-0.204	0.041	0.158	-0.342	-0.028
	(0.217)	(0.178)	(0.235)	(0.210)	(0.225)	(0.249)	(0.187)
Belief on rel. performance (0-1)	-1.129***	-0.393**	0.721**	1.636***	1.366***	1.945***	1.397***
	(0.280)	(0.198)	(0.292)	(0.285)	(0.266)	(0.408)	(0.233)
Overplacement (=1)	0.782***	-0.037	-0.639***	-0.731***	-0.509**	-1.204***	-0.731***
	(0.191)	(0.127)	(0.200)	(0.198)	(0.199)	(0.246)	(0.150)
Patience Index	-0.175**	0.046	0.152*	0.102	0.021	0.173	0.135
	(0.075)	(0.085)	(0.080)	(0.136)	(0.143)	(0.143)	(0.091)
Risk Taking Index	-0.258**	-0.052	0.167	0.324**	0.305**	0.427**	0.304***
	(0.110)	(0.099)	(0.108)	(0.132)	(0.144)	(0.176)	(0.099)
Time Difference between Sessions (in days)							
Time Difference	0.026	-0.020	-0.031	0.000	0.006	-0.013	-0.016
	(0.018)	(0.016)	(0.020)	(0.012)	(0.014)	(0.020)	(0.012)
Constant	-2.312	2.828*	2.916	-4.496***	-4.939***	-1.009	
	(1.990)	(1.551)	(2.149)	(1.076)	(1.432)	(1.754)	
Num.Obs.	950	950	693	950	715	492	950
RMSE	0.40	0.48	0.45	0.39	0.43	0.41	1.85

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

Note: Models (I) - (VI) are logit regressions. Model (I) compares choosing the Fixed (F) over the Variable (V) or Tournament (T) payment. Model (II) compares choosing the Variable over the Fixed or Tournament payment. Model (III) compares choosing the Variable over the Fixed payment. Model (IV) compares choosing the Tournament over the Fixed or Variable payment. Model (V) compares choosing the Tournament over the Variable payment. Model (VI) compares choosing the Tournament over the Fixed payment. Model (VII) is an ordered logit regression (F – V – T). Clustered standard errors on the session level in parentheses.

A.3 List of explanatory variables

Here, we provide a description of the included explanatory variables. All variables were carefully chosen based on their potential to shape earnings and life outcomes as reported in the literature.

Socio-economic status and other relevant socio-demographic variables. SES and demographic variables have been shown to be strongly associated with educational outcomes and earnings (Cunha and Heckman, 2007; Heckman, 2006). Besides age and gender, our questionnaire included a number of proxies for SES,[22] which were mainly informed by three broadly used indexes:

- PISA wealth index: The PISA test provides valuable information to educational researchers and policy makers around the world by comparing countries with regard to their educational system using a variety of educational outcomes. In many ways PISA has emerged as the international benchmark in comparing educational systems (Fuchs and Woessmann, 2008; Hanushek and Woessmann, 2011; Woessmann, 2016). Their SES indicators have often been used for assessing socioeconomic background with teenagers (Hanushek and Woessmann, 2011; West and Woessmann, 2010; Woessmann, 2016). We focus on the family wealth possessions index (WEALTH), which has been validated as a strong and reliable proxy for SES (Rutkowski and Rutkowski, 2013).[23] It includes seven items: *(i)* Do you have a room of your own? Or do you share your room (e.g. with siblings)?; *(ii)* Do you have a link to the Internet at home?; *(iii)* How many cell phones are there at your home?; *(iv)* How many televisions are there at your home?; *(v)* How many computers are there at your home?; *(vi)* How many cars are there at your home?; and *(vii)* How many rooms with a bath or shower are there at your home? In addition to this, we include the number of books available at home, which has been found to alone be another important proxy for socioeconomic status in the PISA test (Woessmann, 2016).

- Family Affluence Scale (FAS) score: This score is also commonly used to elicit SES among school-aged children (Andersen et al., 2008; Boyce et al., 2006; Hartley et al., 2016; Torsheim et al., 2016). The score is similar to the PISA wealth index, and three of the items are the same. It includes four items: *(i)* Do you have a room of your own? Or do you share your room (e.g. with siblings)?; *(ii)* Does your family own a car, van or truck?; *(iii)* How many times did you and your family travel out

[22] Given our sample of adolescents, elicited information about household income must be assumed to be very noisy, which is why we use alternatives.

[23] The questions were drawn from PISA tests conducted in 2015. They were accessed from https://www.oecd-ilibrary.org/education/pisa-2015-assessment-and-analytical-framework/pisa-2015-background-questionnaires_9789264255425-8-en.

of Germany abroad for holiday/vacation last year?; and *(iv)* How many computers does your family own?

- Education and family structure: We follow Kosse et al. (2020) in considering educational and time resources available to the family as important determinants of SES. We thus elicit the following indicators: *(i)* neither parents has a school-leaving degree qualifying for university studies; *(ii)* the participant lives in a single-parent household.[24]

- We collected several other relevant variables: number of siblings, pocket money, migration background (two indicators for being born in Germany, for having German parents) and another migration indicator for speaking a different language than German at home (Hansson and Gustafsson, 2013; Woessmann, 2016)

Aggregation to a composite socio-economic status index using PCA: The items of the above listed three socio-economic indexes are used together with the three migration indicators and the amount of pocket money that the participant receives to create a single component based on principal component analysis, using the weights of the first component. Table A9 summarizes the socio-economic items that load most heavily on this (first) component.

Table A9: Rotated component loading for socio-economic variables

	Socio-economic components
Number of computers	0.387
Number of cars	0.384
Parents German	0.349
Number of bathrooms	0.305

Note: Notes: This table shows the rotated component loading from varimax rotations of principal component analysis of the included fifteen socio-economic variables. Variables with loadings less than 0.3 are excluded from this table.

Reference level of productivity and stress level. A baseline measure of performance was captured in Part 1, where a 5 minutes RET paid on a piece-rate was performed.[25] This serves as a proxy for individual's productivity in playing the real effort task. The reference level of productivity has been found to be important for sorting decisions (Dohmen and Falk, 2011). At the end of the RET (both 5 and 20 minute version), we followed Dohmen and Falk (2011) and elicited self-reported measures of effort, stress, and exhaustion. All

[24] Kosse et al. (2020) consider a third dimension to define SES: household income. As argued above, eliciting actual income with children arguably results in noisy measures, and we thus proxy household income e.g., with the questions included in the PISA wealth index (see above).

[25] Subjects were instructed to solve as many tables as they can, and were paid 0.06 cents for each correctly solved table.

three have been found to be higher in pay for performance schemes compared to a fixed payment.

Beliefs. Prior to starting the 20 minutes real effort task in Part 2, we collected information about a participant's guessed rank in the 5 minute real effort task in Part 1. They got paid according to their guess at the end of the study.[26] Beliefs have been found to be important in sorting decisions, for example in explaining gender differences in sorting into tournaments and differences in productivity (Bordalo et al., 2019; DellaVigna and Pope, 2017; Dohmen and Falk, 2011; Larkin and Leider, 2012; Reuben et al., 2017). Overplacement represents a dummy variable for a positive difference between the belief about own relative performance and actual own relative performance in part 1 (both normalized on the session level).

Cognitive abilities. Cognitive ability has been found to be an important predictor of school attainment as well as future earnings (Borghans et al., 2008; Cawley et al., 2001; Segal, 2012). The main proxy for cognitive ability is the score obtained in the Raven's matrix test administered in Part 1 of the experiment (Raven, 2000). Additionally, we consider self-reported math and German grades. All are expected to be highly correlated with productivity in the real effort task (Dohmen and Falk, 2011).

Altruism, risk, and time preferences. Risk and time preferences predict labor market outcomes, educational attainment, income and wealth (Alan and Ertac, 2018; Becker et al., 2012; Bonin et al., 2007; Cadena and Keys, 2015; DellaVigna and Paserman, 2005; Dohmen et al., 2011; Golsteyn et al., 2014; Sutter et al., 2013; Von Gaudecker et al., 2011). They have also been found to be important for different sorting decisions (Bonin et al., 2007; Dohmen and Falk, 2010, 2011). Altruism, risk, and time preferences are measured by using a subset of the global preference survey by Falk et al. (2018, 2023). For both time and risk preferences, we collected multiple measures: a qualitative measure and a quantitative one (staircase). We combined them as proposed by Falk et al. (2018).

Big five. Personality traits, such as the big five (Openness, conscientiousness, extraversion, agreeableness and neuroticism), have been shown to be stable traits in affecting performance and life outcomes (Almlund et al., 2011; Akee et al., 2018; Cubel et al., 2016; Deming, 2017; Lindqvist and Vestman, 2011; Segal, 2012). We collected data for all big-five, but mostly focus on conscientiousness and neuroticism as they are found to be consistent predictors of performance in various settings (Borghans et al., 2008; Donato et al., 2017; Heckman and Rubinstein, 2001).

[26] See details in the design section.

Competitiveness. A large literature documents gender differences in competitiveness. Women avoid competetive schemes, while men are competing too much (Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2010). These gender differences can potentially explain differences in education and labor market outcomes (Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2010). A high level of competitiveness is also a strong predictor for choosing a more prestigious academic track, controlling for ability (Buser et al., 2014, 2017; Reuben et al., 2017) as well as sorting decisions between different payment schemes (Dohmen and Falk, 2011). Competitiveness is measured in our study on the basis of the Competitive Orientation Measure (one single composite scale; see Newby and Klein, 2014).

Parenting style. Parenting style is important for the academic achievements and future success of children (Doepke et al., 2019; Kosse et al., 2020). We elicited a vital component of parenting style: positive parenting (Essau et al., 2006; Frick, 1991), which indicates the use of positive stimuli and rewards by parents. Recent literature has shown that parental investments have important impact on child cognitive and non cognitive outcomes (Carneiro and Heckman, 2003; Falk et al., 2021; Doepke et al., 2019).

Grit. Grit is defined as perseverance toward a set goal and is seen as being closely related to conscientiousness (Alan et al., 2019). Grit has been found to be predictive of success in a variety of contexts such as through college GPA and educational attainment (Alan et al., 2019; Duckworth et al., 2007). In their study, Alan et al. (2019) finds that students participating in a grit-focused intervention chose more challenging tasks and perform better in the real effort task. It was measured in our study by the short-scale Duckworth Grit Index (Dobbie and Fryer Jr, 2015; Duckworth and Quinn, 2009).

References

Akee, R., W. Copeland, E. J. Costello, and E. Simeonova (2018). How does household income affect child personality traits and behaviors? *American Economic Review 108*(3), 775–827.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Alan, S. and S. Ertac (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy 126*(5), 1865–1911.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 4, Chapter 1, pp. 1–181. Elsevier.

Andersen, A., R. Krolner, C. Currie, L. Dallago, P. Due, M. Richter, Á. Örkényi, and B. Holstein (2008). High agreement on family affluence between children's and parents' reports: International study of 11-year-old children. *Journal of Epidemiology & Community Health 62*(12), 1092–1094.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics 4*(1), 453–478.

Bonin, H., T. Dohmen, A. Falk, D. Huffman, and U. Sunde (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics 14*(6), 926–937.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–773.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Borghans, L., H. Meijers, and B. Ter Weel (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry 46*(1), 2–12.

Boyce, W., T. Torsheim, C. Currie, and A. Zambon (2006). The Family Affluence Scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research 78*(3), 473–487.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review 107*(5), 125–130.

Cadena, B. C. and B. J. Keys (2015). Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy 7*(3), 126–153.

Carneiro, P. M. and J. J. Heckman (2003). Human capital policy. *IZA Discussion Paper* (821).

Cawley, J., J. Heckman, and E. Vytlacil (2001). Three observations on wages and measured cognitive ability. *Labour Economics 8*(4), 419–442.

Cubel, M., A. Nuevo-Chiquero, S. Sanchez-Pages, and M. Vidal-Fernandez (2016). Do personality traits affect productivity? Evidence from the laboratory. *The Economic Journal 126*(592), 654–681.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

DellaVigna, S. and M. D. Paserman (2005). Job search and impatience. *Journal of Labor Economics 23*(3), 527–588.

DellaVigna, S. and D. Pope (2017). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies 85*(2), 1029–1069.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Dobbie, W. and R. G. Fryer Jr (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy 123*(5), 985–1037.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The economics of parenting. *Annual Review of Economics 11*(1), 55–84.

Dohmen, T. and A. Falk (2010). You get what you pay for: Incentives and selection in the education system. *The Economic Journal 120*(546), F256–F271.

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

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Donato, K., G. Miller, M. Mohanan, Y. Truskinovsky, and M. Vera-Hernández (2017). Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in India. *American Economic Review 107*(5), 506–510.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.

Duckworth, A. L. and P. D. Quinn (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment 91*(2), 166–174.

Essau, C. A., S. Sasagawa, and P. J. Frick (2006). Psychometric properties of the Alabama Parenting Questionnaire. *Journal of Child and Family Studies 15*(5), 595–614.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2023). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Management Science 69*(4), 1935–1950.

Falk, A., F. Kosse, P. Pinger, H. Schildberg-Hörisch, and T. Deckers (2021). Socioeconomic status and inequalities in children's IQ and economic preferences. *Journal of Political Economy 129*(9), 2504–2545.

Frick, P. J. (1991). The Alabama Parenting Questionnaire. *University of Alabama*.

Fuchs, T. and L. Woessmann (2008). What accounts for international differences in student performance? A re-examination using PISA data. In C. Dustmann, B. Fitzenberger, and S. Machin (Eds.), *The Economics of Education and Training*, pp. 209–240. Physica-Verlag HD.

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

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014). Adolescent time preferences predict lifetime outcomes. *The Economic Journal 124*(580), F739–F761.

Hansson, Å. and J.-E. Gustafsson (2013). Measurement invariance of socioeconomic status across migrational background. *Scandinavian Journal of Educational Research 57*(2), 148–166.

Hanushek, E. A. and L. Woessmann (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3, Chapter 2, pp. 89–200. Elsevier.

Hartley, J. E., K. Levin, and C. Currie (2016). A new version of the HBSC Family Affluence Scale - FAS III: Scottish qualitative findings from the international FAS development study. *Child Indicators Research 9*(1), 233–245.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science 312*(5782), 1900–1902.

Heckman, J. J. and Y. Rubinstein (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review 91*(2), 145–149.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). The formation of prosociality: Causal evidence on the role of social environment. *Journal of Political Economy 128*(2), 434–467.

Larkin, I. and S. Leider (2012). Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *American Economic Journal: Microeconomics 4*(2), 184–214.

Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–128.

Newby, J. L. and R. G. Klein (2014). Competitiveness reconceptualized: Psychometric development of the competitiveness orientation measure as a unified measure of trait competitiveness. *The Psychological Record 64*(4), 879–895.

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

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in Math test scores: The role of competition. *Journal of Economic Perspectives 24*(2), 129–144.

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology 41*(1), 1–48.

Reuben, E., M. Wiswall, and B. Zafar (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.

Rutkowski, D. and L. Rutkowski (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education 8*(3), 259–278.

Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science 58*(8), 1438–1457.

Sutter, M., M. G. Kocher, D. Gältzle-Rützler, and S. T. Trautmann (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review 103*(1), 510–531.

Torsheim, T., F. Cavallo, K. A. Levin, C. Schnohr, J. Mazur, B. Niclasen, and C. Currie (2016). Psychometric validation of the revised Family Affluence Scale: A latent variable approach. *Child Indicators Research 9*(3), 771–784.

Von Gaudecker, H.-M., A. Van Soest, and E. Wengstrom (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review 101*(2), 664–694.

West, M. R. and L. Woessmann (2010). 'Every Catholic child in a Catholic school': Historical resistance to state schooling, contemporary private competition and student achievement across countries. *The Economic Journal 120*(546), F229–F255.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives 30*(3), 3–32.

A.4 Instructions

Choice & Exogenous Treatments

Instructions for Part 1

Create your ID

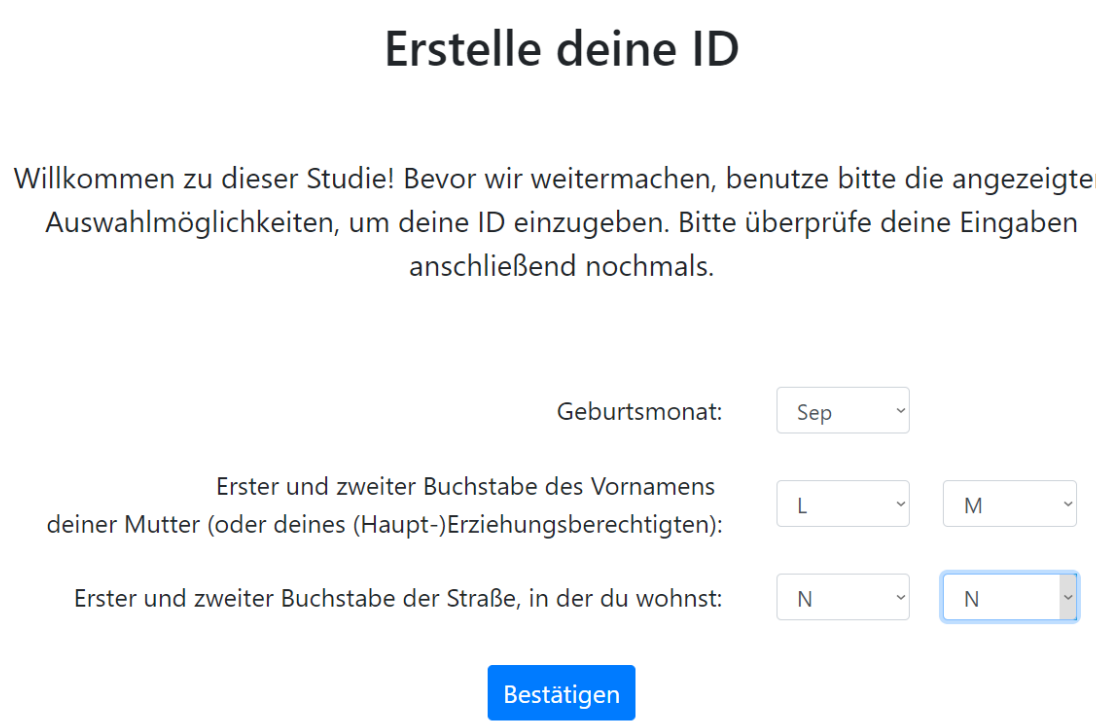
Experimenter reads aloud: *[Welcome to the study. This study consists of two sessions: this session today, and another session in which you will participate in the near future. In both sessions you will earn money, please listen carefully to the instructions.* ***For today, you will receive a fixed payment of 4€ if you complete the session. You can also earn additional money depending on your performance in a task that I will explain later.*** *During the session you cannot talk to the other students in this room. This is a very important rule, and if you break it, you will not receive the money that you earned. On the first page, you will be asked to enter your ID as explained on the screen. Your name will never be used during the entire study. Whatever you do, and all the answers you give will only be recorded under your ID. That means that everything you do in the study is going to be anonymous. When you will participate in the second session, you will also be identified via the same anonymous ID code. If you have any questions, please raise you hand and one of us will come to your desk to answer it in private. Please remember that your participation on this study is fully voluntary, and you can decide to quit at any time. If you decide to quit before finishing the study, you are not allowed to leave the room, and you are still required to stay seated at your desk.]*

Welcome to this study. Before we proceed, use the drop-downs to enter your ID. Please double check all your entries before proceeding, as it is very important that your ID is specified correctly.

- Month of birth *[drop down – Jan, Feb, Mar, ...]*
- First and second letter of your mother's first name (or your legal guardian's first name)
- First and second letter of the street where you live

CONFIRM

$\Longrightarrow$ ——— *Enter ID (first trial)* ——— $\Longleftarrow$



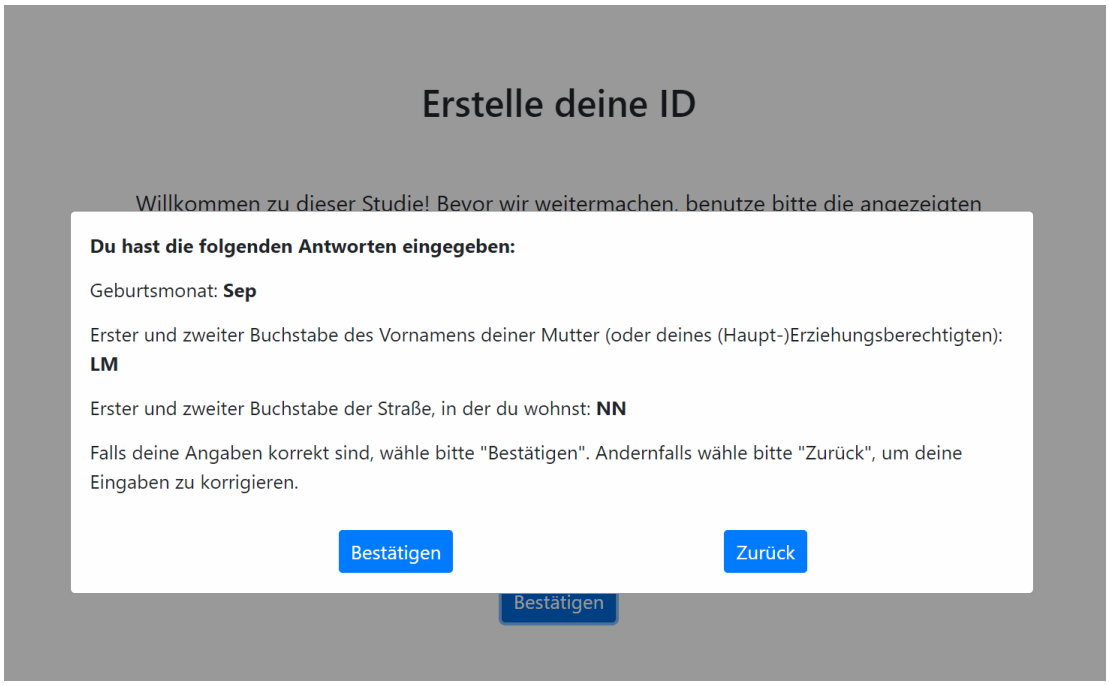
$\Longrightarrow$ ——— *pop-up* ——— $\Longleftarrow$

You provided the following answers:

- Month of birth: _____
- First and second letter of your mother's first name (or your legal guardian's first name): _____
- First and second letter of the street where you live: _____

If your answer is correct please press CONFIRM otherwise press BACK to revise your entries.

CONFIRM BACK



$\Longrightarrow$ ——— *wait for all & new screen for double IDs* ——— $\Longleftarrow$

Your ID is the same of someone else in this room. We hence ask you to answer an additional question:

- Last two letter of your first name

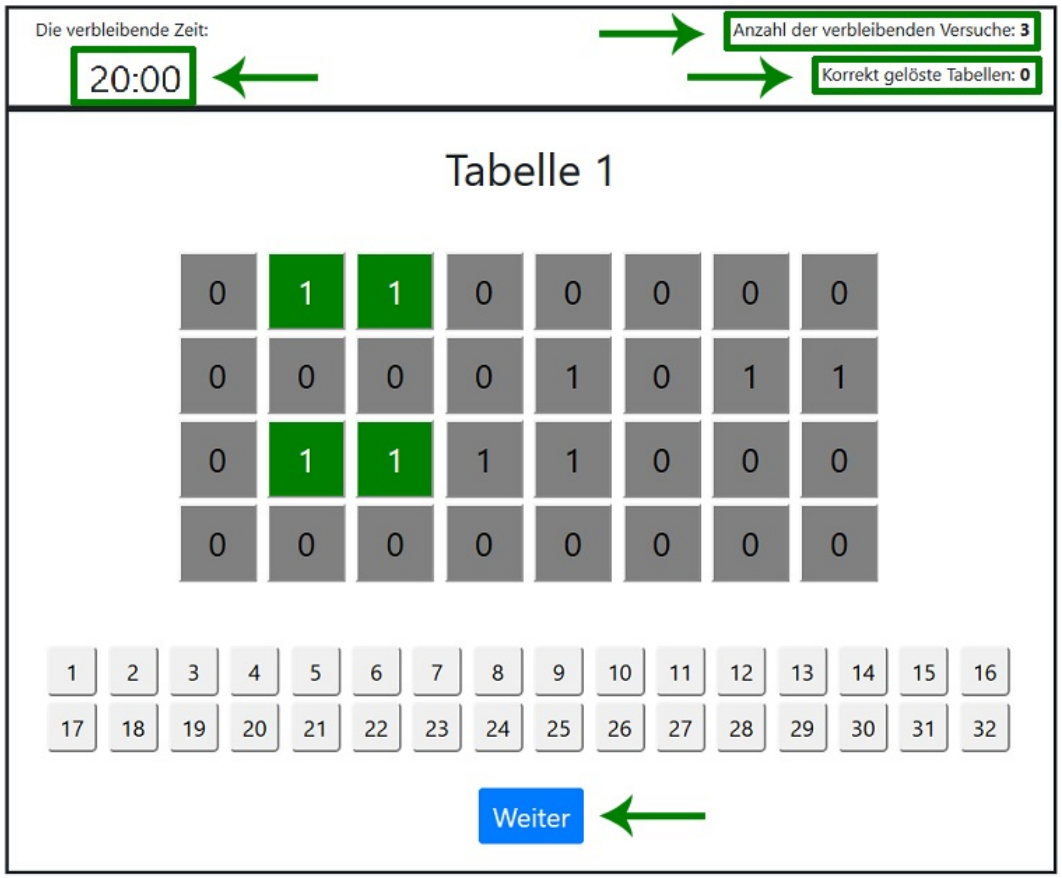
CONFIRM

$\Longrightarrow$ ——— *wait for all & new screen* ——— $\Longleftarrow$

Your Task

Experimenter reads aloud:

You will be shown a set of tables with 1s and 0s, like the one reported below. Your task is to correctly solve as many tables as you can.



To correctly solve a table, you have to:

1. **Tap** on all the cells containing a 1, which will highlight them in a dark green color
2. **Count** the correct amount of 1s that you see in the table, and report this amount in the number pad underneath the table.

Be aware, you are **not** allowed to highlight the 0s! If you accidentally highlight a 0, you can tap on the cell again to change it back to grey.

Once you are done with the tapping and you have reported the number, press CONFIRM. You get three tries to solve a table correctly. You will see the amount of remaining tries in the upper-right corner. If you do not manage to solve a table within the three tries, the next table will be shown on your screen. There are no penalties for not solving a table. You can see the amount of correctly solved tables in the upper right corner at any point during the task.

You have a total of 5 minutes to solve as many tables as you can. You will be **paid 0.06€ for each table you solved correctly.** For instance, if by the end of the 5 minutes you solve 1 table correctly, you will earn 0.06€. If by the end of the 5 minutes you solve 10 tables correctly, you will earn 10 times 0.06€, so you will earn 0.6€. Or for instance, if by the end of the 5 minutes you solve 100 tables correctly, you will earn 100 times 0.06€, so you will earn 6€.

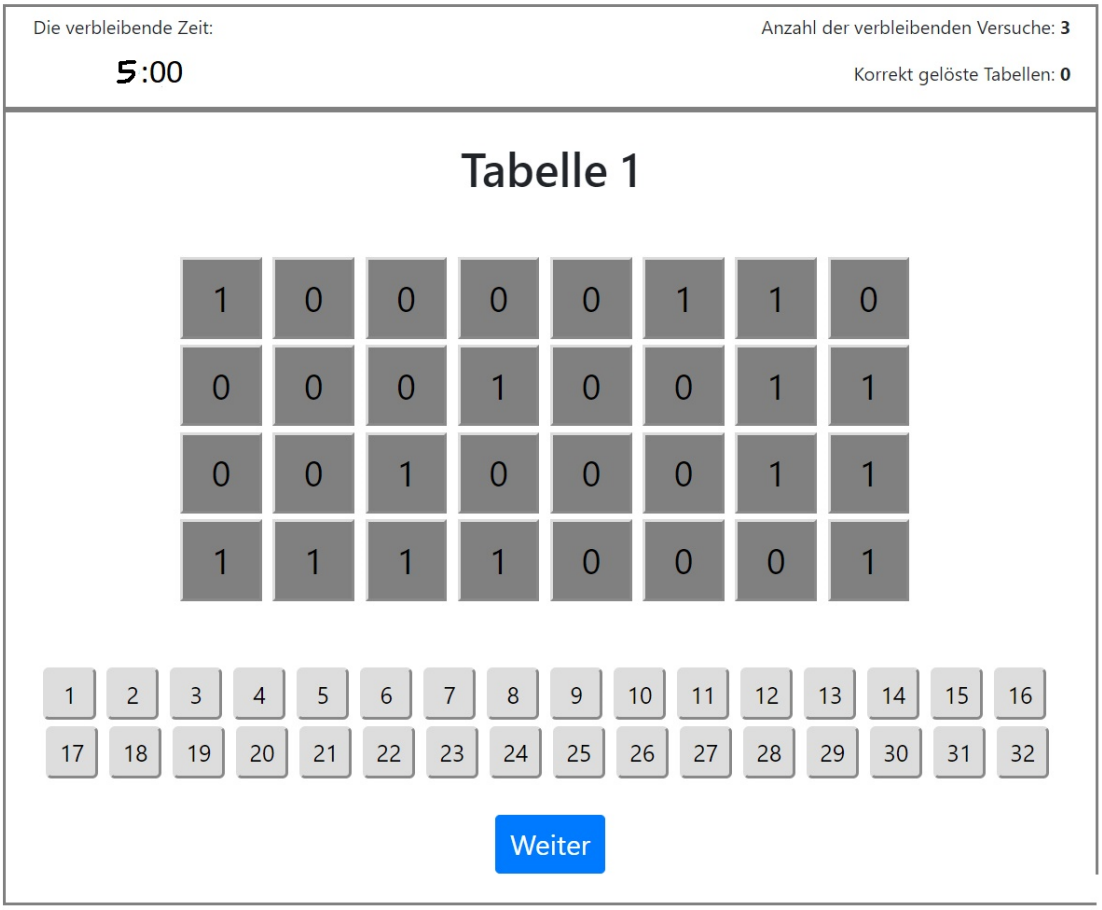
Before you start with the task you will have one trial round. That means that the first table you solve will not count for money, but will help you get acquainted with the task. After you correctly solve the first table, the 5 minute period will start.

Remember that you are not allowed to talk to the other participants in this room. If you have any question, please rise you hand and one of us will come to your desk to answer it privately. [27]

[27] The program is advanced by the experimenter after about 2-3 min (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue the experiment.

$\Longrightarrow$ ——— *wait for all & new screen*[28] ——— $\Longleftarrow$

The real effort task



[28] A similar table with "Trial round" is displayed. The picture is the same without the remaining time and correctly solved tables. After subjects correctly solve the trial round, they enter a waiting screen which lasts until everybody solves the trial round. Then a new screen appears with a 5 second countdown: "The 5 minute period for solving the task will start in 5, 4, 3... " Following that, Table 1 is displayed and the 5 minute period starts.

$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

The following questions are related to the task you completed. Please answer the following questions referring to the task you just solved. Please indicate your answers on a 7 point scale, where 1 means “not at all" and 7 means “very much":

	1	2	3	4	5	6	7
How much effort did you exert?	○	○	○	○	○	○	○
How stressed did you feel?	○	○	○	○	○	○	○
How exhausted did you get?	○	○	○	○	○	○	○

CONTINUE

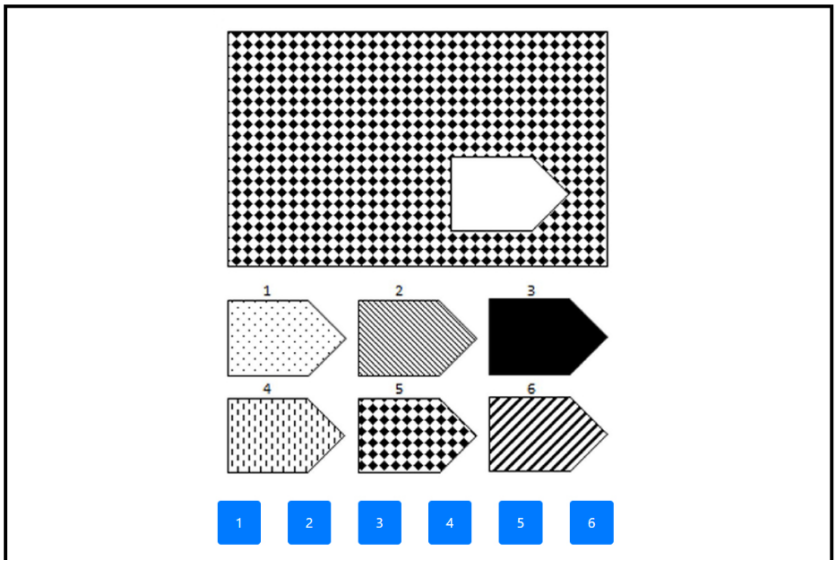
⟹ ——— *new screen* ——— ⟸

For the following tasks you have to look at the picture, and find the missing piece of the picture. Once you find it, you need to circle it, as it is shown in the example below. Your goal is to solve as many tasks within 5 minutes as possible. If you have any questions, please raise your hand. If everything is clear, you can start immediately.[29]

Rätselaufgabe

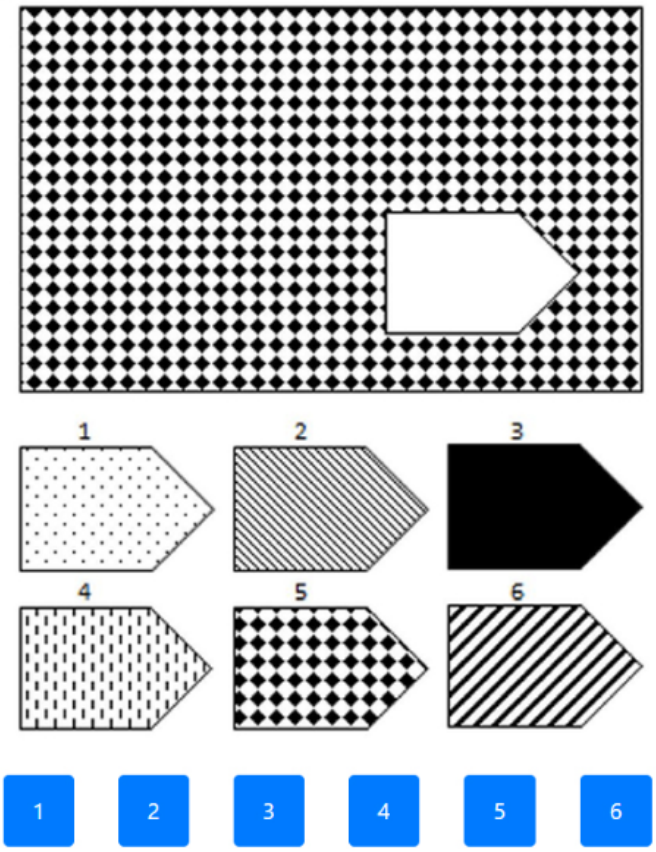
In der nächsten Aufgabe werden ihnen einige Bilder wie unten abgebildet gezeigt. Sie müssen sich das jeweils gezeigte Bild genau anschauen und das fehlende Teil finden, welches das Bild vervollständigt.

Sobald sie das fehlende Stück gefunden haben, tippen sie auf die entsprechende Nummer im unteren Teil des Bildschirms und bestätigen ihre Auswahl. Ihr Ziel ist es, möglichst viele Bilder innerhalb von 5 Minuten korrekt zu lösen. Sollten sie eine Frage haben, heben sie bitte ihre Hand und ein Mitarbeiter wird zur Klärung zu ihnen kommen.



NEXT

[29] Instructions are given with an example to make the task clear to the participants. The program is then advanced by the experimenter after they are done with reading the instructions (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue with the task.

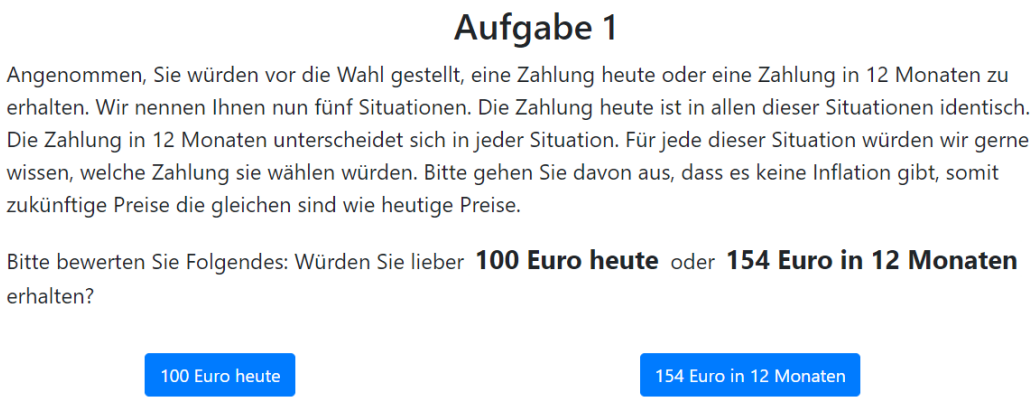


[30]The matrices are shown and the 5 minutes count-down starts to solve the 10 different matrices. Participants cannot go back once they have submitted an answer for a given task or skip between pictures. If a participant completes the tasks before the 5 minutes, they would have to wait for the other participants to finish.

[31]When subjects have answered all the matrices in the IQ task, they are told to proceed with the remainder of the tasks by themselves.

$\Longrightarrow$ ——— *new screen* – ***staircase for time preferences*** ——— $\Longleftarrow$

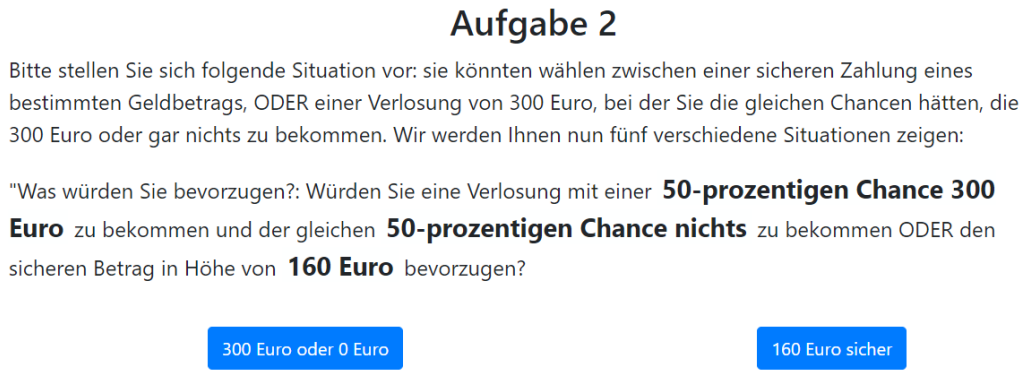
Suppose you were given the choice between the following: receiving a payment today or a payment in 12 months. We will now present to you five situations. The payment today is the same in each of these situations. The payment in 12 months is different in every situation. For each of these situations we would like to know which you would choose.[32]



[32]The two different options are shown with the corresponding amounts and for all 5 pages and the different amounts are highlighted to make them salient. Subjects advance to the next page when they click on one of the two buttons. The staircase approach is taken from Falk et al. (2023, 2018).

$\Longrightarrow$ ——— *new screen* – ***staircase for risk preferences*** ——— $\Longleftarrow$

Please imagine the following situation: You can choose between a sure payment and a lottery. The lottery gives you a 50 percent chance of receiving 300 €. With an equally high chance you receive nothing. Now imagine you had to choose between the lottery and a sure payment. We will present to you five different situations. The lottery is the same in all situations. The sure payment is different in every situation.[33]



[33]The two different options are shown with the corresponding amounts and for all 5 pages and the different amount are highlighted to make them salient. Subjects advance to the next page when they click on one of the two buttons. The staircase approach is taken from Falk et al. (2023, 2018)

⟹ ——— *new screen* ——— ⟸

⟹ ——— ***Questionnaire*** ——— ⟸

1. Are you female or male?	O Female	O Male	O I don't want to comment
2. What is your zip code?			
3. Where are you born?	O In Germany O In another EU country O In a European country outside of EU O In an Asian country O In an African country O In a South American country O In a North American country O In Australia		
4. What grade are you in?	O Grade 10 O Grade 11 O Grade 12 O Grade 13		
5. Year of birth?			
6. Grade in math? (final grade for last school year)	O 1 O 2 O 3 O 4 O 5 O 6		
7. Grade in German? (final grade for last school year)	O 1 O 2 O 3 O 4 O 5 O 6		
8. If everything goes as planned, when do you plan to finish the Abitur? (If you don't plan to finish the Abitur, please answer "No plans about finishing the Abitur")	O 2019 O 2020 O 2021 O 2022 O 2023 O No plans about finishing the Abitur		
9. How much pocket money/allowance do you get per week?	0-95 € per week		
10. Do you have a mother/father born outside of Germany?	O Both parents born in Germany O Mother born outside of Germany O Father born outside of Germany O Both parents born outside Germany		

11. Do you live together with one or two parents (legal guardians)? (If you live with one parent and his/her partner, please answer: Two parents)	O One parent	O Two parents	O Neither
12. What is the highest education level of your mother?	O University or similar O High school O Middle school or lower O No schooling O I don't know		
13. What is the highest education level of your father?	O University or similar O High school O Middle school or lower O No schooling O I don't know		
14. What do you plan to do after you finish high school?	O University degree in STEM (Science, Technology, Engineering and Mathematics) O University degree outside of STEM O Vocational training (Ausbildung) O I want to find a job O I want to take some time off O Voluntary military service O I don't know		
15. Do you have any siblings?	O 0 O 1 O 2 O 3 O 4 or more		
16. How many books are there in your home?	O 0-10 books O 11-25 books O 26-100 books O 101-200 books O 201-500 books O More than 500 books		
17. What languages do you speak at home most of the time?	O German O English O Turkish O Spanish O Italian O French O Arabic O Other		

18. How many times did you and your family travel out of Germany abroad for holiday/vacation last year?	O None O Once O Twice O More than twice		
Which of the following are in your home?			
19. A room of your own?	O Yes	O No	
20. A link to the Internet?	O Yes	O No	
How many of these are there at your home?			
21. Cell phones?	O None O One O Two O Three or more		
22. Televisions?	O None O One O Two O Three or more		
23. Computers/PCs?	O None O One O Two O Three or more		
24. Cars?	O None O One O Two O Three or more		
25. Rooms with a bath or shower?	O None O One O Two O Three or more		

$\Longrightarrow$ ——— *new screen – from preference module* ——— $\Longleftarrow$

Please tell me, in general, how willing or unwilling you are to take risks. Please use a scale from 0 to 10, where 0 means you are "completely unwilling to take risks" and a 10 means you are "very willing to take risks". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

0 = Completely unwilling to take risks										*Very willing to take risks = 10*
○	○	○	○	○	○	○	○	○	○	○
0	1	2	3	4	5	6	7	8	9	10

We now ask for your willingness to act in a certain way in four different areas. Please again indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so". You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

	Completely **unwilling** to do so										Completely **willing** to do so
How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?	○	○	○	○	○	○	○	○	○	○	○
	0	1	2	3	4	5	6	7	8	9	10

$\Longrightarrow$ ——— *new screen – BFI-44* ——— $\Longleftarrow$

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please select a number next to each statement to indicate the extent to which you agree or disagree with that statement.

I see Myself as Someone Who...	1. Disagree strongly	2. Disagree a little	3. Neither agree nor disagree	4. Agree a little	5. Agree strongly
1. Is talkative	○	○	○	○	○
2. Tends to find fault with others	○	○	○	○	○
3. Does a thorough job	○	○	○	○	○
4. Is depressed, blue	○	○	○	○	○
5. Is original, comes up with new ideas	○	○	○	○	○
6. Is reserved	○	○	○	○	○
7. Is helpful and unselfish with others	○	○	○	○	○
8. Can be somewhat careless	○	○	○	○	○
9. Is relaxed, handles stress well	○	○	○	○	○
10. Is curious about many different things	○	○	○	○	○
11. Is full of energy	○	○	○	○	○
12. Starts quarrels with others	○	○	○	○	○
13. Is a reliable worker	○	○	○	○	○
14. Can be tense	○	○	○	○	○
15. Is ingenious, a deep thinker	○	○	○	○	○
16. Generates a lot of enthusiasm	○	○	○	○	○
17. Has a forgiving nature	○	○	○	○	○
18. Tends to be disorganized	○	○	○	○	○
19. Worries a lot	○	○	○	○	○
20. Has an active imagination	○	○	○	○	○
21. Tends to be quiet	○	○	○	○	○

I see Myself as Someone Who...	1. Disagree strongly	2. Disagree a little	3. Neither agree nor disagree	4. Agree a little	5. Agree strongly
23. Tends to be lazy	○	○	○	○	○
24. Is emotionally stable, not easily upset	○	○	○	○	○
25. Is inventive	○	○	○	○	○
26. Has an assertive personality	○	○	○	○	○
27. Can be cold and aloof	○	○	○	○	○
28. Perseveres until the task is finished	○	○	○	○	○
29. Can be moody	○	○	○	○	○
30. Values artistic, aesthetic experiences	○	○	○	○	○
31. Is sometimes shy, inhibited	○	○	○	○	○
32. Is considerate and kind to almost everyone	○	○	○	○	○
33. Does things efficiently	○	○	○	○	○
34. Remains calm in tense situations	○	○	○	○	○
35. Prefers work that is routine	○	○	○	○	○
36. Is outgoing, sociable	○	○	○	○	○
37. Is sometimes rude to others	○	○	○	○	○
38. Makes plans, & follows through with them	○	○	○	○	○
39. Gets nervous easily	○	○	○	○	○
40. Likes to reflect, play with ideas	○	○	○	○	○
41. Has few artistic interests	○	○	○	○	○
42. Likes to cooperate with others	○	○	○	○	○
43. Is easily distracted	○	○	○	○	○
44. Is sophisticated in art, music, or literature	○	○	○	○	○

$\Longrightarrow$ ——— *new screen* – ***Revised Competitiveness Index*** ——— $\Longleftarrow$

The following scale measures aspects of competitiveness. Please read each question carefully and try to answer as honestly as possible. Do not spend too much time on any one item; if trying to decide between two responses, choose the one that first comes to mind.

	1. Strongly disagree	2. Slightly disagree	3. Neither agree nor disagree	4. Slightly agree	5. Strongly agree
1. I like competition.	○	○	○	○	○
2. I am a competitive individual.	○	○	○	○	○
3. I enjoy competing against an opponent.	○	○	○	○	○
4. I don't like competing against other people.	○	○	○	○	○
5. I get satisfaction from competing with others.	○	○	○	○	○
6. I find competitive situations unpleasant.	○	○	○	○	○
7. I dread competing against other people.	○	○	○	○	○
8. I try to avoid competing with others.	○	○	○	○	○
9. I often try to outperform others.	○	○	○	○	○
10. I try to avoid arguments.	○	○	○	○	○
11. I will do almost anything to avoid an argument.	○	○	○	○	○
12. I often remain quiet rather than risk hurting another person.	○	○	○	○	○
13. I don't enjoy challenging others even when I think they are wrong.	○	○	○	○	○
14. In general, I will go along with the group rather than create conflict.	○	○	○	○	○

$\Longrightarrow$ ——— *new screen* – ***Alabama Parenting Style (positive parenting)*** ——— $\Longleftarrow$

The following are statements about your family. Please rate each item and how often it TYPICALLY occurs in your home.

		1. Never	2. Almost Never	3. Sometimes	4. Often	5. Always
1	Your parents tells you that you are doing a good job.	○	○	○	○	○
2	Your parents reward you or give you something extra to you for behaving well.	○	○	○	○	○
3	Your parents compliment yuo when you have done something well.	○	○	○	○	○
4	Your parents prise you for behaving well.	○	○	○	○	○
5	Your parents hug or kiss you when you done something well.	○	○	○	○	○
6	Your parents tell you that they like it when you help out around the house.	○	○	○	○	○

$\Longrightarrow$ ——— *new screen* – ***short grit scale*** ——— $\Longleftarrow$

Please respond to the following 8 items. Be honest – there are no right or wrong answers!

	1. Not like me at all	2. Not much like me	3. Somewhat like me	4. Mostly like me	5. Very much like me
1. New ideas and projects sometimes distract me from previous ones.	○	○	○	○	○
2. Setbacks don't discourage me.	○	○	○	○	○
3. I have been obsessed with a certain idea or project for a short time but later lost interest.	○	○	○	○	○
4. I am a hard worker.	○	○	○	○	○
5. I often set a goal but later choose to pursue a different one.	○	○	○	○	○
6. I have difficulty maintaining my focus on projects that take more than a few months to complete.	○	○	○	○	○
7. I finish whatever I begin.	○	○	○	○	○
8. I am diligent	○	○	○	○	○

$\Longrightarrow$ ——— *new screen at the end of part 1* ——— $\Longleftarrow$

Thanks for taking part in the study.

In the first task you solved _____ tables correctly.

You earnings for this task are: _____€ (rounded up at the 10 cents)

In addition, you earned a 4€ fee for taking part in the study.

Your total earnings for today are: _____€

Please remain seated and remember that you are not allowed to talk to the other participants. One of the experimenters will come to your desk to give you your earnings.

Choice Treatment[34]
Instructions for Part 2[35]

Create your ID

Experimenter reads aloud: *[In this study you will earn money, so please listen carefully to the instructions. During the study you cannot talk to the other students in this room. This is a very important rule, and if you break it, you will not receive the money that you earned during the study. On the first page, you will be asked to enter your ID as explained on the screen. Your name will never be used during the study. Whatever you do, and all the answers you give will only be recorded under your ID. That means that everything you do in the study is going to be anonymous! If you have any questions, please raise you hand and one of us will come to your desk to answer it in private. Please remember that your participation on this study is fully voluntary, and you can decide to quit at any time. If you decide to quit before finishing the study, you are not allowed to leave the room, and you are still required to stay seated at your desk.]*

Welcome to this study! Before we proceed, use the drop-downs to enter your ID. Please double check all your entries before proceeding, as it is very important that your ID is specified correctly.

- Month of birth *[drop down – Jan, Feb, Mar, ...]*
- First and second letter of your mother's first name (or your legal guardian's first name)
- First and second letter of the street where you live
- Last two letter of your first name[36]

CONFIRM

[34] Same instructions as for exogenous, except the subjects face no choice screen, and only information about the relevant payment scheme is diplayed.
[35] The original German instructions and available upon request from the authors.
[36] Extra question in case of double ID.

$\Longrightarrow$ ——— *Enter ID (first trial)* ——— $\Longleftarrow$

Erstelle deine ID

Willkommen zu dieser Studie! Bevor wir weitermachen, benutze bitte die angezeigten Auswahlmöglichkeiten, um deine ID einzugeben. Bitte überprüfe deine Eingaben anschließend nochmals.

Geburtsmonat:	Sep	
Erster und zweiter Buchstabe des Vornamens deiner Mutter (oder deines (Haupt-)Erziehungsberechtigten):	L	M
Erster und zweiter Buchstabe der Straße, in der du wohnst:	N	N

Bestätigen

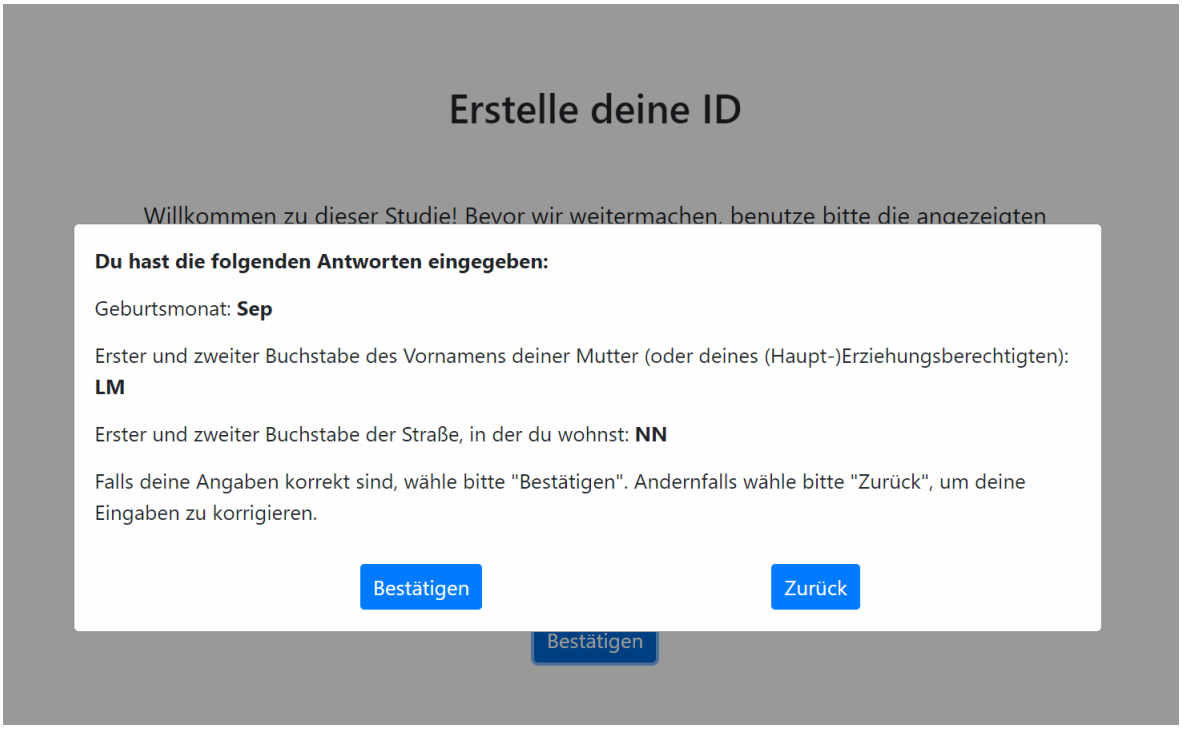
$\Longrightarrow$ ——— *pop-up* ——— $\Longleftarrow$

You provided the following answers:

- Month of birth: _____
- First and second letter of your mother's first name (or your legal guardian's first name): _____
- First and second letter of the street where you live: _____

If your answer is correct please press CONFIRM otherwise press BACK to revise your entries.

CONFIRM BACK



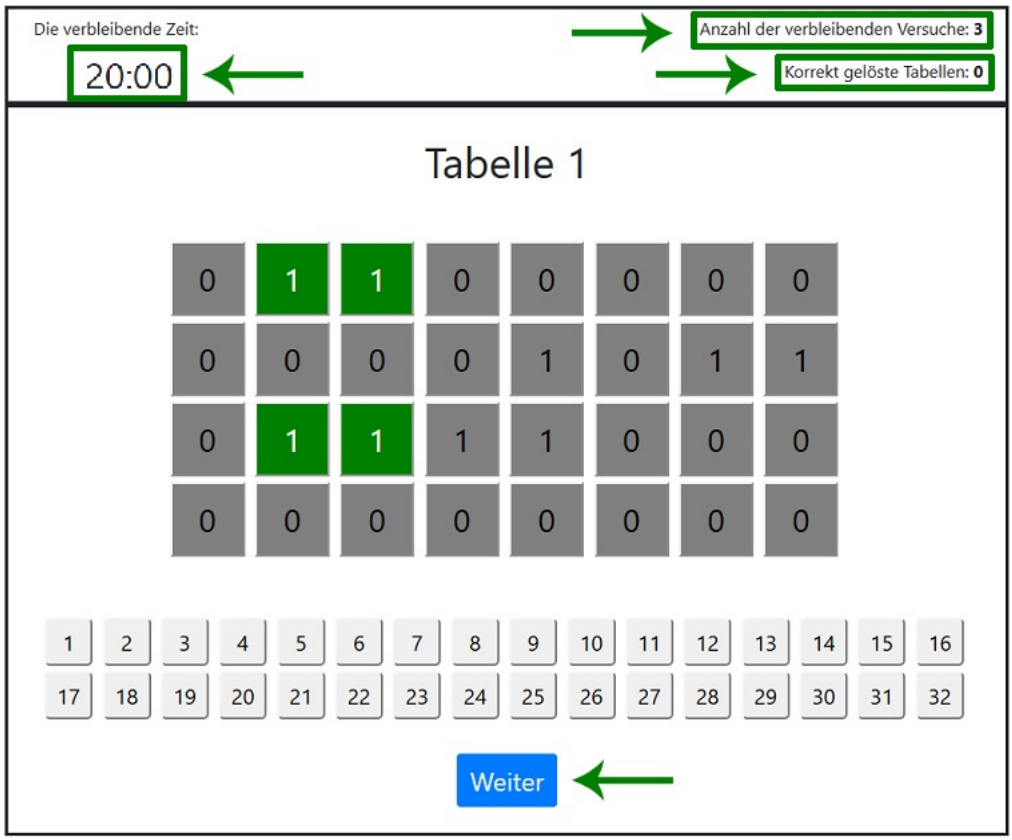
$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

$\Longrightarrow$ ——— *wait for all & new screen* ——— $\Longleftarrow$

Your Task

Experimenter reads aloud: *[I will now explain you the task in which you can earn money. Some of you have already seen the task as you did it the first time we came to your class. But some of you were not here; to be certain that you all know the task, I will explain it in detail again. Please follow the instructions carefully.]*

You will be shown a set of tables with 1s and 0s, like the one reported below. Your task is to correctly solve as many tables as you can.



To correctly solve a table, you have to:

1. **Tap** on all the cells containing a 1, which will highlight them in a dark green color;
2. **Count** the correct amount of 1s that you see in the table, and report this amount in the number pad underneath the table.

Be aware, you are **not** allowed to highlight the 0s! If you accidentally highlight a 0, you can tap on the cell again to change it back to grey.

Once you are done with the tapping and you have reported the number, press CONFIRM. You get three tries to solve a table correctly. You will see the amount of remaining tries in the upper-right corner. If you do not manage to solve a table within the three tries, the next table will be shown on your screen. There are no penalties for not solving a table. You can see the amount of correctly solved tables in the upper right corner at any point during the task. **You have a total of 20 minutes for solving the task.**

Remember that you are not allowed to talk to the other participants in this room. If you have any question, please raise you hand and one of us will come to your desk to answer it privately.[37]

[37]The program is advanced by the experimenter after about 2-3 min (A "continue" button is displayed for the subjects once the experimenter advances the program) and participants are told to click "Continue" once they are ready to continue the experiment

$\Longrightarrow$ ——— *new screen*[38] ——— $\Longleftarrow$

Guessing Task

Before explaining how you will be paid for the task, you have a chance to earn some additional money.

In this room, there are XY students (including you) that were present also during our previous visit. You all performed the task for 5 minutes the other time.

In the 5-minute version of the task, **you correctly solved X tables.**

We ranked you and the other participants present in the previous visit.[39] You were all ranked based on the number of tables correctly solved in 5 minutes. For example, position number 1 is for the one who solved the most tables, position number 2 is for the one who solved the second most tables, and so on, with the last position XY for the one who solved the least.

We would now like you to guess your position in the ranking.

If you were to guess the correct number, you earn 2€. If you come within up to 5 positions (higher or lower), you will earn 0.50€ that will be added to your total earnings for today's session.

\- ——————————————————————————— +

1 | | | | | | | | | | | | | | | | | XY

I think I ranked number .....

CONFIRM[40]

[38] Screen only appears for subjects that are present in both sessions.

[39] Participants that are present in both sessions are ranked by standard competition ranking.

[40] Participants need to touch the slider to activate it. They can adjust the number either by touching the slider or clicking on the + and - signs at the ends of the slider. Absolute numbers of the different options for the ranking appear after the participant click on the slider. "I think I ranked number ...." only appears when participants click on the slider with the number of the ranking clicked on.

$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

Your Earnings[41]

You can determine the payment mode yourself. In particular, you can choose between three alternative payment modes.

Fixed Payment. When the 20 minutes are up, you will receive 6.5€, independent of the number of tables you solved correctly.

Variable Payment. When the 20 minutes are up, you will be paid 0.06€ for each table you solved correctly.

Tournament. When the 20 minutes are up, you will be paid either 0.08€ or 0.04€ *for each table* you solved correctly. To establish whether you will be paid 0.08€ or 0.04€ per correct table, your performance will be compared with one other student in this room, whose payment will also be determined in the same way. At the end of the 20 minutes, if you solved more tables than this other student matched with you, you will get 0.08€ per correct table. If instead you solved less tables than this other participant matched with you, you will get 0.04€ per correct table. If you and this other participant solved the same number of tables, the computer will randomly determine if you are paid 0.08€ or 0.04€ per correct table.

For your information, in the first visit you have solved _____ in _____ minutes. [42]

Before choosing your payment mode, please answer a few control questions designed to make sure you understood how the earnings are computed. If you have any questions, please raise your hand and wait for an experimenter to come to your desk. [43]

1. In the **fixed payment** alternative, if you solve 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.60€

   b I will get 6.50€

   c I will get 80.00€

2. In the **fixed payment** alternative, if you solve 1000 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.60€

[41] In the exogenous treatment, subjects would only be able to see the paragraph explaining the payment scheme that they were assigned, and would only receive control questions referring to that payment scheme.

[42] This info was displayed only if the ID was present in first study and it is unique in second study.

[43] The correct answers are marked here in bold for display.

b I will get 6.50€

c I will get 80.00€

3. In the **variable payment** alternative, if you solve 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.60€

   b I will get 6.50€

   c I will get 90.00€

4. In the **variable payment** alternative, if you solve 1000 tables correctly by the end of 20 minutes, how many Euros will you get?

   a I will get 0.60€

   b I will get 6.50€

   c I will get 60.00€

5. In the **tournament payment** alternative, if you solve 1000 tables correctly, and the student matched with you solves 10 tables correctly by the end of the 20 minutes, how many Euros will you get?

   a I will get 0.40€

   b I will get 6.50€

   c I will get 80.00€

6. In the **tournament payment** alternative, if you solve 10 tables correctly, and the student matched with you solves 1000 tables correctly by the end of the 20 minutes, how many Euros will you get?

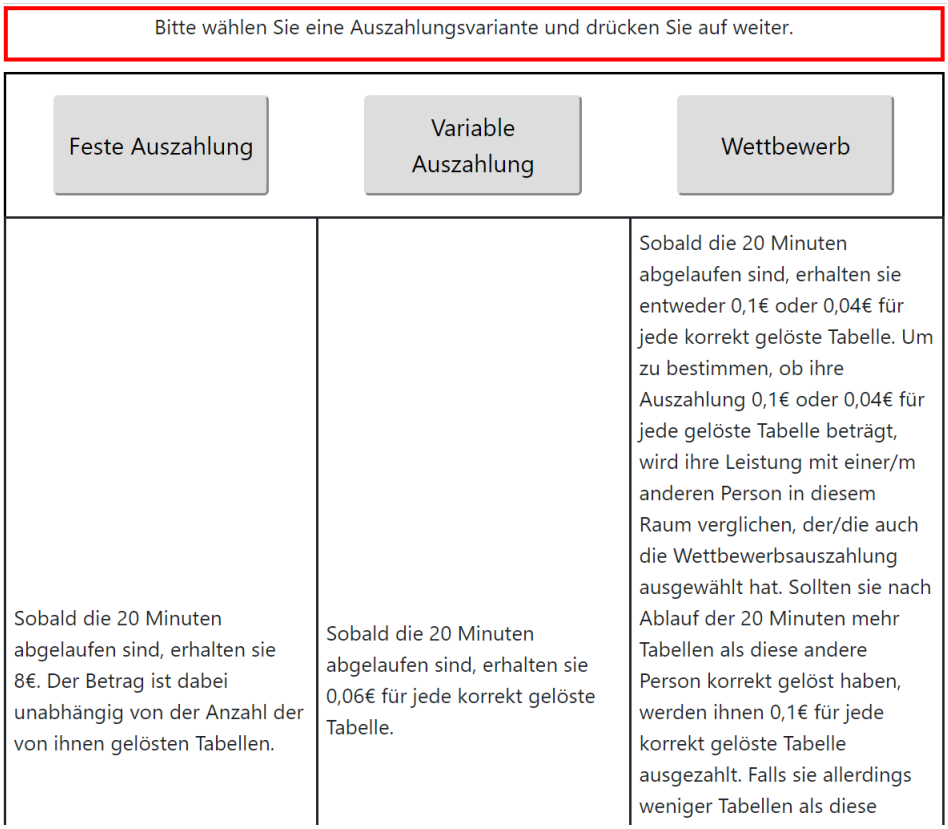
   a I will get 0.40€

   b I will get 6.50€

   c I will get 80.00€

⟹ ——— *new screen* ——— ⟸

Choice of Payment Mode



⟹ ——— *wait for all & new screen* ——— ⟸

Show if only one person chose tournament

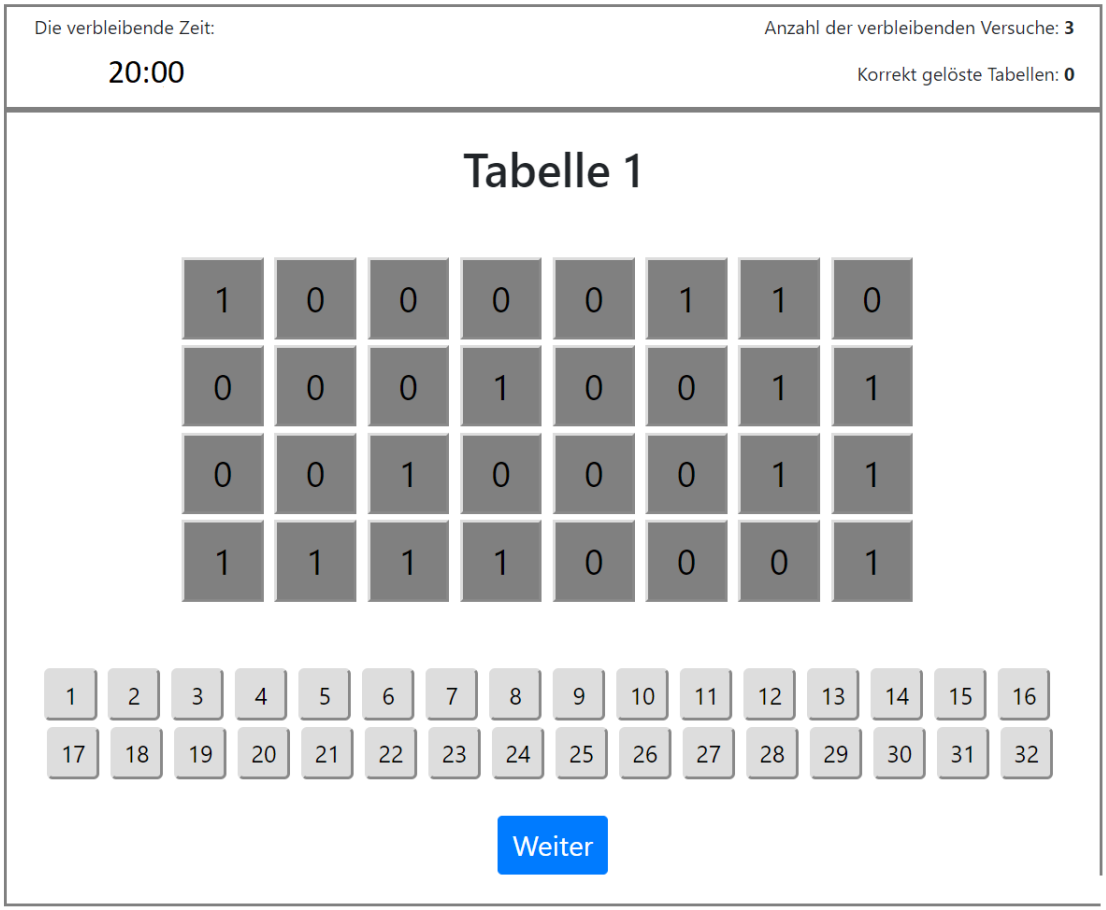
You are the only one who chose tournament. Unfortunately, it is not possible to match you with another student in this room. Please choose again, this time between fixed and variable payment.[44]

⟹ ——— *wait for all & new screen*——— ⟸

Countdown. The task will start in 10, 9, 8...seconds.

[44] A menu with the two possible choices are shown to the participant.

$\Longrightarrow$ ——— *new screen – the real effort task* ——— $\Longleftarrow$



$\Longrightarrow$ ——— *new screen* ——— $\Longleftarrow$

Before proceeding to the final payments, please answer the following questions referring to the task you just solved. Please indicate your answers on a 7 point scale, where 1 means “not at all" and 7 means “very much":

	1	2	3	4	5	6	7
How much effort did you exert?	○	○	○	○	○	○	○
How stressed did you feel?	○	○	○	○	○	○	○
How exhausted did you get?	○	○	○	○	○	○	○

NEXT

$\Longrightarrow$ ——— *new screen at the end of the task* ——— $\Longleftarrow$

Thank you for taking part in our study. The task is now over.

You solved _____ tables correctly.

You earnings for this task are: _____€

In addition, you earned a 1€ fee for taking part in the study.

In the guessing task you earned ..... €

[**Only for Tournament.**] You solved more/less/the same number of tables than/as the student you are compared to.
[**Only in case of tie.**] The computer randomly determined that you earn 0.08/0.04€ per solved table.

Your earnings for the task are: _____ €
(rounded up at the 10 cents)

Please remain seated and remember that you are not allowed to talk to the other participants. One of the experimenters will come to your desk to give you your earnings.