10.14754/CEU.2022.08

Essays on the Behavioral Economics of Inequality

by
Luca Flóra Drucker

Submitted to
Central European University
Department of Economics and Business

In partial fulfilment of the requirements for the degree
Doctor of Philosophy in Economics

Supervisor: Botond Kőszegi
Co-advisor: Dániel Horn

Budapest, Hungary

2022

DOI:10.14754/CEU.2022.08

CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF ECONOMICS AND BUSINESS

The undersigned hereby certify that they have read and recommend to the Department of Economics and Business for acceptance a thesis entitled **"Essays on the Behavioral Economics of Inequality"** by Luca Flóra Drucker.
Dated: September 26, 2022

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Chair of the Thesis Committee:

DocuSigned by:
A. Weber
____
Andrea Weber

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Advisor:

DocuSigned by:
Botond Kőszegi
____
Botond Kőszegi

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Co-Advisor:

DocuSigned by:
____
Dániel Horn

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
Internal Examiner:

DocuSigned by:
Mats Köster
____
Mats Köster

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.
External Examiner:

DocuSigned by:
Alexander Cappelen
____
Alexander Cappelen

DOI:10.14754/CEU.2022.08

I certify that I have read this dissertation and in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

External Member:

DocuSigned by:

Gergely Hajdu

301B8F1F20CD4C3...

Gergely Hajdu

10.14754/CEU.2022.08

Copyright notice

Author: Luca Flóra Drucker
Title: Essays on the Behavioral Economics of Inequality
Degree: Ph.D.
Dated: October, 2022

Hereby I testify that this thesis contains no material accepted for any other degree in any other institution and that it contains no material previously written and/or published by another person except where appropriate acknowledgement is made.

Signature of the author: ... Drucker Luca ...

10.14754/CEU.2022.08

Co-author contribution

Chapter 2: Excuse-driven Present Bias

Joint work with Marc Kaufmann

The paper was developed in close cooperation with Marc Kaufmann throughout all stages. We contributed equally to the idea of the paper, data management, programming, analysis and writing.

10.14754/CEU.2022.08

Abstracts

This thesis consists of three chapters on different behavioral aspects of inequality. In the first, single-authored chapter, I run an online experiment to look at whether people compensate for inequalities in difficulty when rewarding performance. In the second chapter, co-authored with Marc Kaufmann, we use two online experiments to study whether people postpone more work to the future when they have excuses. In the third, single-authored chapter I look at how children from different socioeconomic backgrounds adjust their educational aspirations after having to repeat a grade at the end of primary school.

Chapter 1: Difficult Merits

There is evidence that the average person accepts income inequality that is based on differences in individual achievement. However, a particular achievement is not equally difficult for everyone. I study how information about differences in difficulties affects redistributive decisions. I consider two sources of difficulties: external circumstances and individual ability. Participants have to redistribute the income earned by achievement within pairs with information about the relative difficulty of the task within the pair. I find that participants strongly compensate the member of the pair who had a harder job in producing if the difficulties come from an external source, but fully ignore the relative difficulties when they arise from individual ability. This is true for both when participants redistribute between themselves and another subject and when they redistribute between two other subjects. Nevertheless, when involved, participants choose allocations that benefit them the most: those with externally harder tasks make more selfish, more egalitarian, and more compensating allocation choices, while those with easier tasks are more likely to simply reward performance, even though they learned that performance differences were caused partly by external differences in the difficulty of the task.

Chapter 2: Excuse-driven Present Bias

Joint work with Marc Kaufmann

We test whether people behave in a more present-biased way when they can excuse such behavior. We run two experiments, one on the Amazon Mechanical Turk and one with students in Luxembourg, to elicit subjects' willingness to work (WTW) today and at a future date. We elicit this WTW against an alternative that provides no excuses and one that provides an excuse. In the first experiment, while the no-excuse alternative always requires participants to work harder in the future, the excuse alternative adds a 10% chance of future work remaining easy. We find that the WTW today drops by \$0.11 more than the WTW in two days when we move from the no-excuse to the excuse alternative, as if the excuse alternative is worth

10.14754/CEU.2022.08

more when it allowed postponing hard work to the future. This result cannot be explained by risk and time preferences that do not depend on other alternatives present. In the second experiment, we test the excuse of a chance of not having to do extra work in the future, and another potential excuse: a different type of task in the future. The results do not support that a different task would act as an excuse for postponing work. For the chance of no extra work, we get non-significant results that nevertheless point in the same direction as the MTurk results. We discuss both experiments and describe a planned follow-up study with the goal of replicating our findings with excuses based on risk.

Chapter 3: Compensatory Advantage and Inequality in Educational Aspirations

There is a large body of evidence supporting that family background determines the development of children from early on. Children from higher and lower socioeconomic backgrounds enter the education system with different cognitive and non-cognitive skills. Besides differences in early childhood, a child's socioeconomic status (SES) can also affect how she copes with hardships at later stages in her educational career. Using two large, administrative and survey-based, datasets from Hungary, I look at how children from different socioeconomic backgrounds change their educational aspirations after one specific hardship – grade retention in the 7th grade of primary school. Using the difference-in-differences method I find that children from all socioeconomic backgrounds decrease their aspirations after retention, but the magnitudes are larger for low-SES children. The post-retention SES gap in aspirations is the highest for those children who had high aspirations before retention. However, when looking at subsamples of children by 6th-grade mathematics performance, the effects are heterogeneous. For those children who failed in mathematics at the mid-term of 6th grade – so those who most likely repeated 7th grade due to their low performance at school –, there is no compensatory advantage of high socioeconomic background either in 8th-grade aspirations or in transitioning to a secondary school that gives access to tertiary education. For children who had better grades in mathematics in their 6th year, there is, on the other hand, a compensatory advantage in both outcomes.

10.14754/CEU.2022.08

Acknowledgements

I am indebted to my advisors, Botond Kőszegi and Dániel Horn, for guiding me through my PhD and giving me constant support and regular feedback on my work. They also helped me find opportunities to learn more at summer schools and to present my work at conferences for which I am deeply grateful.

I thank Dániel Horn and Hubert János Kiss for inviting me to participate in a project about measuring school children's non-cognitive skills in their classrooms. I learned a lot about designing and running experiments through this project, and it was also this project that inspired me to do experiments.

I am grateful for Marc Kaufmann for being interested in my thoughts about procrastination and for asking me to write a paper together on excuse-driven present bias. I have learned invaluable academic and technical skills through working with Marc on this project.

I also thank Bertil Tungodden and Alexander Cappelen for advising me during my online visit at FAIR and for involving me in the rich academic life at the Research Center, despite the fact that I could not travel there because of Covid restrictions.

I thank my two examiners, Alexander Cappelen and Mats Köster, who gave thorough and constructive feedback that helped me improve my papers substantially.

I thank Miklós Koren, Arieda Muço and Andrea Weber for starting the Empirical Research course, which helped me narrow down the ideas I was thinking about at that time to a question that lead to the third chapter of this thesis.

I am also grateful for other faculty members at CEU whom I could always turn to with my questions and concerns and who gave valuable comments and suggestions regarding my projects. I also thank the staff members at the department who were always very helpful.

I am thankful for my fellow PhD students who helped me through all stages of the PhD: we wrote homework together, participated in trainings and summer schools, spent endless hours in the lab doing our research, attended the gym and had long conversations over coffee in the department kitchen.

Finally, I thank my family and my partner who have been constantly supporting me and who have encouraged me to pursue my dreams.

Contents

1 Chapter 1: Difficult Merits 1
1.1 Introduction . . . 1
1.2 Experiment details . . . 3
1.2.1 Task length treatment . . . 4
1.2.1.1 Production . . . 4
1.2.1.2 Redistribution . . . 4
1.2.2 Ability treatment . . . 5
1.2.2.1 Production . . . 5
1.2.2.2 Redistribution . . . 7
1.2.3 Participants . . . 8
1.3 Descriptive statistics . . . 9
1.3.1 Production . . . 9
1.3.1.1 Ability measure . . . 10
1.3.2 Redistribution . . . 10
1.4 Reduced-form analysis . . . 13
1.5 Structural analysis . . . 18
1.5.1 Descriptive model . . . 18
1.5.2 Structural analysis . . . 20
1.6 Conclusion . . . 24

2 Chapter 2: Excuse-driven Present Bias 26
2.1 Introduction . . . 26
2.2 Identification of excuse-driven present bias . . . 28
2.3 First experiment . . . 30
2.3.1 Technical details . . . 30
2.3.1.1 Payments . . . 31
2.3.2 Timeline . . . 31
2.3.2.1 The Tutorial . . . 31
2.3.2.2 The Main Sessions . . . 32
2.3.2.3 Choosing large matrices rather than extra matrices . 33
2.3.3 Implementation of risk as an excuse . . . 34
2.4 First experiment results . . . 34
2.4.1 Main results . . . 35
2.4.2 Ruling out concavity of money . . . 35
2.5 Second experiment . . . 37
2.5.1 A chance of no extra tasks as an excuse . . . 37
2.5.2 A different type of task as an excuse . . . 37
2.5.3 Technical details . . . 38
2.6 Second experiment results . . . 39
2.7 Discussion . . . 43

3 Chapter 3: Compensatory Advantage and Inequality in Educational Aspirations **45**
3.1 Introduction . . . 45
3.2 Data . . . 47
3.2.1 National Assessment of Basic Competencies data . . . 47
3.2.1.1 Background questionnaire . . . 48
3.2.2 Administrative database . . . 48
3.2.2.1 Health care data . . . 48
3.2.3 Definition of variables . . . 49
3.2.3.1 Grade retention rules in Hungary . . . 50
3.3 Descriptive statistics . . . 50
3.3.1 Aspirations . . . 50
3.3.2 Retention . . . 52
3.3.3 8^{th} grade aspirations and secondary school tracks . . . 56
3.3.4 Missing data . . . 57
3.4 Baseline empirical analysis . . . 58
3.4.1 8^{th} grade aspirations . . . 58
3.5 Subsamples of repeaters by 6^{th} grade mathematics performance . . . 62
3.6 10^{th} grade outcomes . . . 65
3.7 Discussion . . . 68

A Appendix for Chapter 1 **75**
A.1 Design . . . 75
A.2 Distribution of tasks/minute within group . . . 76
A.3 Production . . . 77
A.4 Redistribution . . . 78
A.5 Attrition . . . 79
A.6 Robustness checks of reduced-form results . . . 80
A.6.1 Sample with all demographic controls . . . 80
A.7 Experiment instructions . . . 84
A.7.1 Production . . . 84
A.7.2 Redistribution – task length spectators . . . 90
A.7.3 Redistribution – task length stakeholders . . . 92
A.7.4 Redistribution – ability spectators . . . 94
A.7.5 Redistribution – ability stakeholders . . . 96

B Appendix for Chapter 2 **98**
B.1 Choices in batches 1 and 2 of the first experiment . . . 98
B.2 Debrief survey statistics . . . 98
B.3 Choices in the second experiment . . . 100
B.3.1 Risk batches . . . 100
B.3.2 Matrix-versus-Greek batches . . . 101

B.3.3 Follow-up risk . . . 103
B.3.4 Ruling out reference dependence . . . 104
B.4 First experiment instructions . . . 106
B.4.1 Session 0 . . . 106
B.4.2 Session 1 . . . 113
B.4.3 Session 2 . . . 118
B.5 Second experiment instructions . . . 119
B.5.1 Session 1 each week . . . 125
B.5.2 Session 2 each week . . . 130

C Appendix for Chapter 3 **132**
C.0.1 Descriptive tables . . . 132
C.0.2 Retention probabilities . . . 133
C.0.3 Robustness checks for main regressions . . . 135
C.0.3.1 Using three categories of SES . . . 135
C.0.3.2 Using continuous SES . . . 137
C.0.3.3 Regression of continuous aspirations on the Admin3 data . . . 140
C.0.3.4 Regressions on subsamples with health controls . . . 141
C.0.4 Data Appendix . . . 147
C.0.4.1 Birth year and month and school starting age variables 147
C.0.4.2 Creating years of education from the categorical education variables . . . 148
C.0.4.3 Admin3 NABC extension . . . 148

List of Tables

1.1 Demographic characteristics of participants 8
1.2 Mean tasks/minute by group in each treatment 9
1.3 Time taken for the 10 tasks in the ability treatment 11
1.4 Spectator decisions in the task length treatment 15
1.5 Spectator decisions in the ability treatment 16
1.6 Stakeholder decisions in the task length treatment 17
1.7 Stakeholder decisions in the task length treatment 18
2.1 Timeline of the first experiment 31
2.2 Results: t-tests 36
2.3 T-test equivalents, ruling out concavity of money 36
2.4 Timeline of the second experiment 39
2.5 Main results 40
2.6 Short-term $\Delta\Delta$ by preferred/disliked task 40
2.7 Follow-up risk choices 41
2.8 Reduced-form direction of $\beta\delta$ and value of the 10% drop 42
2.9 $\Delta\Delta$ 42
2.10 Short-term $\Delta\Delta$ by type of task 43
2.11 Final survey results 44
3.1 Educational aspirations of 6^{th} grade children by socioeconomic status 51
3.2 Repeating 7th grade grade by socioeconomic status 52
3.3 Summary statistics of low- and high-SES 7th grade repeaters 54
3.4 Regression of retention on SES and various characteristics. 55
3.5 Regression of retention on SES and various characteristics by 6th year midterm mathematics grade. 56
3.6 Probability of 8^{th} grade aspirations being at most vocational education 59
3.7 Probability of 8^{th} grade aspirations being tertiary education 60
3.8 Educational aspirations in 8^{th} grade in years 61
3.9 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm 63
3.10 8^{th} grade aspirations for those only passing mathematics at 6^{th} grade midterm 64
3.11 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm 66
3.12 Attending a secondary school ending with a high school diploma (1 in mathematics) 67
3.13 Attending a secondary school ending with a high school diploma (2 in mathematics) 68
3.14 Attending a secondary school ending with a high school diploma (3-5 in mathematics) 69
A.1 Production by group and self-reported effort level 77
A.2 Attrition from first to second part 79

A.3 Spectator decisions in the task length treatment 80
A.4 Spectator decisions in the ability treatment 81
A.5 Stakeholder decisions in the task length treatment 82
A.6 Stakeholder decisions in the task length treatment 83
B.1 Summary statistics from debrief survey 99
C.1 Observations and share of repeaters in each aspiration-SES cell 132
C.2 Change in the educational aspirations in years by parental education 132
C.3 Characteristics of repeaters by 6^{th} year mathematics grades and SES: failing (1) and just passing (2) mathematics 133
C.4 Characteristics of repeaters by 6^{th} year mathematics grades and SES: grades 3-5 134
C.5 Probability of 8^{th} grade aspirations being at most vocational education 135
C.6 Probability of 8^{th} grade aspirations being tertiary education 136
C.7 Probability of 8^{th} grade aspirations being at most vocational education 137
C.8 Probability of 8^{th} grade aspirations being at most vocational education 138
C.9 Educational aspirations in 8^{th} grade in years 139
C.10 Educational aspirations in 8^{th} grade in years, Admin3 data 140
C.11 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm, Admin3 database 141
C.12 8^{th} grade aspirations for those just passing mathematics at 6^{th} grade midterm, Admin3 database 142
C.13 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm, Admin3 database 143
C.14 Attending a secondary school ending with a high school diploma (1 in mathematics, Admin3 database) 144
C.15 Attending a secondary school ending with a high school diploma (2 in mathematics, Admin3 database) 145
C.16 Attending a secondary school ending with a high school diploma (3-5 in mathematics, Admin3 database) 146

List of Figures

1.1 Example 3-letter task . . . 5
1.2 Example of a redistribution decision screen . . . 6
1.3 Mean self-reported effort level by group . . . 7
1.4 Group mean production by treatment . . . 10
1.5 Share allocated to a random participant in the pair (spectators) / to self (stakeholders) . . . 12
1.6 Excess income share allocated by situation (spectator decisions) . . . 13
1.7 Excess income share allocated by situation (stakeholder decisions) . . . 14
1.8 Shares of fairness types among spectators . . . 22
1.9 Shares of fairness types among stakeholders . . . 22
1.10 Shares of fairness types among stakeholders by task length . . . 23
1.11 Shares of fairness types among stakeholders by ability . . . 23
1.12 Shares of fairness types among spectators by own task length . . . 24
1.13 Shares of fairness types among spectators by own ability . . . 25
2.1 Example of a large matrix . . . 32
2.2 Example of a price list . . . 33
2.3 Excuse-driven difference in WTW in two days over WTW today . . . 35
2.4 Example of a Greek transcription task . . . 38
3.1 Percentage of retained students in higher grades of primary school . . . 51
3.2 Change in educational aspirations between 6th and 8th grade by SES and repeating . . . 52
3.3 Probability of repeating 7th grade by 6th year midterm mathematics grade . . . 53
3.4 Probability of attending a type of secondary school by 8^{th} grade aspirations . . . 57
A.1 Screenshot of the survey at the end of the experiment . . . 75
A.2 Distribution of tasks/minute within group . . . 76
A.3 Self-reported effort level in each stage of the first part . . . 77
A.4 Excess income share allocated to participants by group (spectator decisions) . . . 78
A.5 Excess income share allocated to self by group (stakeholder decisions) 78
C.1 Probability of repeating 6th grade by 6th year midterm mathematics grade . . . 133

10.14754/CEU.2022.08

1 Chapter 1: Difficult Merits

1.1 Introduction

A stylized fact emerging from the enormous experimental-economics literature on distributional preferences is that people want to reduce income inequality if it is based on pure luck, but not if it reflects differences in merit (e.g., Almås et al., 2010; Durante et al., 2014; Cappelen et al., 2017). This is true even if the inequality is large and only part of it is due to merit (Cappelen et al., 2017), and if individuals had different incentives to perform well (Andre, 2021). An important reality of life, however, is that achieving a given level of performance is transparently more difficult for some than for others. A student who has her own quiet room for studying and is not constantly interrupted by family members, for example, can in the same amount of time prepare better for her upcoming exam than a student who is crammed into a small corner with her siblings running around.[1] At an even more basic level, a student who is more naturally gifted in the subject can learn more in the same amount of time. The existing literature does not study the implications of individuals' knowledge about such advantages for distributional social preferences.

In this paper I study in an experimental setting the extent to which people reduce income inequality reflecting inequality in performance that is partly determined by inequality in difficulties. Numerous experiments established that people are heterogeneous in the redistribution they prefer given the sources of inequality, but the majority of them wants to condition the allocation on factors they can influence and not on factors they cannot (Konow, 2000; Cappelen et al., 2013; Durante et al., 2014). In my experiment I distinguish between two types of difficulty which participants cannot influence. First, subjects, like the students with different home environments above, face randomly assigned differences in how long it takes to achieve a given performance. In addition, just like with studying, subjects may arrive at the task with different abilities. While arguably both differences are exogenous, at least for the type of task I use, I find that subjects treat them very differently, seemingly drawing the line of responsibility not between exogenous and endogenous, but between external and internal factors in performance. While they compensate the person with exogenously harder tasks, they leave the inequality due to differences in ability unchanged. This is true both for the case when subjects redistribute income between themselves and another subject, and the case when they redistribute between two other subjects.

There is evidence that when making decisions over their own income, people frequently use the fairness principle that benefits them the most. This is true both when people are paid unequally ex post (Rodriguez-Lara and Moreno-Garrido, 2012),

[1] It is actually well established that home environment plays a significant role in the development of cognitive and non-cognitive skills, and affects later life outcomes as well (Bradley et al., 2000; Mott, 2004).

and when they faced unequal opportunities ex ante (Eisenkopf et al., 2013; Deffains et al., 2016; Fehr and Vollmann, 2020). Deffains et al. (2016) and Fehr and Vollmann (2020) find that people hold a self-serving bias regarding the source of their success: successful people attribute their success to hard work, while unsuccessful people think they are so because they were unlucky, even though success was exclusively determined by luck – receiving harder or easier tasks – in the experiments. Fehr and Vollmann (2020) also find that successful participants are less likely to want to learn the true source of their success than unsuccessful participants. I add to the evidence on self-serving fairness principles by showing that, with external difficulties, even with full information about the source of their and their partner's success, subjects act on the fairness views that benefit them the most.

The experiment is structured as follows: there is a production and a redistribution part. In the production part, participants are doing a series of simple letter-encryption tasks that take 10-15 seconds to do each. I first obtain a measure of how much time each subject takes for a task with maximal effort by asking subjects to do 10 tasks as fast as they can. The time they need gives a good measure of the difficulty of the task. After the 10 tasks, they have to perform similar tasks again, but now for 15 minutes. This is the actual production, and subjects receive 10 experimental tokens for each completed task within the time frame. The payment for a given production is therefore the same, but for some the tasks take more time to complete than for others. The source of the differences in the time needed for a task can be either external or internal: in the task length treatment, participants face one of three task lengths – long, medium, or short – randomly. This causes external variation in the time needed for a task. Participants then do the 10 tasks and work for 15 minutes with their assigned task length. In the other, ability treatment, all participants have medium-length tasks. Therefore, when measuring how fast subjects complete 10 tasks, differences in the time needed for a task come from an internal source: subjects' differential ability in the task.

In the second, redistribution part, participants within a treatment are randomly assigned into pairs, and they decide how to redistribute either their and their partner's joint income – as *stakeholders* – or another pair of subjects' income – as *spectators* –, with knowledge about both subjects' production and difficulty of the task. In the task length treatment, the difficulty is the length of the task, and the number of tasks subjects with similar task length could do within a minute on average. In the ability treatment it is the ability of the subject in the task – low, medium, or high –, and the number of tasks subjects with similar ability could do within a minute. I find that in the task length treatment, while also conditioning their allocation on the relative production of the participants, spectators give around 4 percentage points more of the joint income to those who had longer tasks than their partners. There is no such compensation for difficulties in the ability treatment: controlling for their production share, participants get the same share of the joint income independent of their relative ability within the pair. Stakeholders in the task length treatment give themselves a 2.5-4.7 percentage points higher share of the joint income if they had longer tasks than their partner. Stakeholders with shorter tasks, however, do not give themselves significantly more or less than with equal task lengths. Again, there is no such pattern in the ability treatment. If anything, subjects with higher ability than their partner give themselves 2.9-3.3 percentage points more than with equal abilities. However, this difference is not robust to different specifications.

I also compute the shares of people with different fairness preference types in a structural framework. The three types that I can distinguish in the experiment are the meritocratic – who allocates the income based on performance –, the egalitarian – who always divides the income equally –, and a compensating type – who rewards performance but compensates the ones with a more difficult job in performing. The results support the reduced-form results: in the task length treatment, 53 percent of spectators are classified as having meritocratic preferences, 17 percent are egalitarians, and 29 percent compensators. In the ability treatment, on the other hand, 83 percent of spectators are meritocrats, 14 percent are egalitarians, and there are virtually zero compensators. The picture is similar for stakeholders: in the task length treatment, 51 percent is classified as meritocrats, 20 percent are egalitarians, and 19 percent are compensators. 9 percent of task length stakeholders are making fully selfish decisions, meaning that they allocate all of the tokens to themselves in most decisions. Long-task stakeholders, however, are less likely to make meritocratic decisions, and more likely to make all other types of decisions than stakeholders with shorter tasks. In the ability treatment, 69 percent of stakeholders are meritocrats, 21 percent are egalitarians, 6 percent are fully selfish, and virtually zero are compensators.

In the next section I explain the experiment design in detail. In Section 3.3 I show descriptive statistics about the production and the redistributive decisions of participants. Section 1.4 presents a reduced-form analysis of the redistributive decisions. Section 1.5 first introduces the model on which I base my structural analysis, and then the structural analysis and results. In Section 1.6 I discuss the results and conclude.

1.2 Experiment details

The experiment took place entirely online. I used the oTree software for coding (Chen et al., 2016a), and Prolific for recruiting participants. I chose participants currently living in the United States. The experimental design, the hypotheses, and the main empirical analysis were pre-registered in the AEA RCT Registry (Drucker, 2021). Due to the interactive nature of the experiment, I chose to do it in two parts, on two consecutive days. The two parts together took around 35 minutes to complete. Participants did the production part on the first day, and the next day they could come back to the redistribution part anytime between 6 AM and midnight. Separating the two parts enabled subjects not having to sit in front of the computer at the same time, excluding attrition due to having to wait for other subjects' move. It also reduced the potential effect of exhaustion after or emotions evoked by the production part on redistributive decisions. On the other hand, it introduced attrition between the two parts. Attrition, however, was not systematic to any experimental or participant characteristics (see Section 1.3.2 for more detail). The experiment took place between 1-9 December, 2021, in three two-day sessions, with 100-250 participants per session. At the beginning of the experiment, participants were assigned into either the task length treatment or the ability treatment. I explain the two treatments in more detail below.

1.2.1 Task length treatment

1.2.1.1 Production First, participants learned the description of the experiment and tried the task. The task was developed by Benndorf et al. (2018) and it consists of encrypting letter combinations into numbers. See an example 3-letter task in Figure 1.1. The letter-number pairs and the order of the pairs in the encryption key are randomized between each correctly solved task, to minimize learning in the task. The fact that there is minimal scope for learning excludes differential learning abilities affecting production. In the task length treatment, every participant was randomly assigned a task length: they faced 2, 3, or 4-letter tasks. After reading the description of the experiment and completing a few practice tasks, participants arrived at a consent page where they decided if they wish to participate in the experiment. If they gave consent, next they had to do 10 tasks with their assigned length as fast as they can. At this point, participants did not know yet that other participants had different task lengths. From the time it took the participants to do the 10 tasks, I calculated how many tasks they could do within a minute. The average tasks/minute within a task length group therefore showed the *actual difficulty* of a task with a particular length. After this stage, participants completed the Production stage, where they worked on tasks of their assigned length for 15 minutes. They received 10 experimental tokens for each correctly solved task. Participants were aware from the beginning of this stage that this was not their final income, but it were to be redistributed in the second part by either themselves or another subject.

1.2.1.2 Redistribution In the second part, participants were randomized into pairs and learned a more detailed description of this part. The pairs were first assigned into either a *spectator* or a *stakeholder* role, and they learned the instructions for their subsequent choices accordingly. Spectators redistributed the joint income of two other subjects, while stakeholders redistributed the joint income of themselves and their partner. They made the redistribution decisions with strategy method: they faced ten decisions about ten random pairs of participants from the first part, but only one of these was the true pairing that had a chance to be implemented at the end. Participants knew this but did not know which pair was the true one. They

10.14754/CEU.2022.08

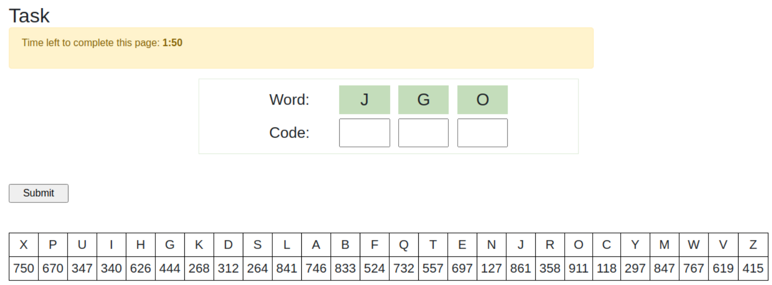


Figure 1.1: Example 3-letter task

Note: Benndorf et al. (2018) task. In the example task, the letter J corresponds to the number 861, G to 444, and O to 911, so the participant has to enter these three numbers into the boxes.

made all redistribution decisions knowing the production levels and the task lengths – short, medium, or long – of both members of the pair, and the mean tasks/minute people could do in the particular task length group.

I chose to show the mean difficulty of a task length instead of each participant's own tasks/minute measure to separate the external difficulty of the task from how the actual person can perform that task. This way when redistributing, decision-makers who care about the difficulties in producing compare the production of the participants to how difficult their tasks were, and not to how many tasks the particular participant could have done had she exerted maximal effort. This also means that, compared to the average difficulty of the task of a specific length, subjects' production depend not only on their effort but also on their ability to perform the task.

Figure 1.2 shows an example of a decision screen. Participants had to make the decisions using a slider that determined the tokens given to each participant in the pair. The participant with lower production in the pair was always on the left-hand side, under the name Participant 1, and the high-performer participant was Participant 2. In the case of stakeholders, the decision-maker participant was always called 'you' in the decision, and her data were either on the left or on the right side, depending on whether her production was below or above the production of her partner in that particular pair. At the end of the experiment, one decision out of the two made for a pair – by either the two members of the pair or the two members of another pair – was chosen randomly and implemented. I converted the experimental tokens to British Pounds, which is the currency participants are paid in on Prolific.

1.2.2 Ability treatment

1.2.2.1 Production The structure of the Ability treatment was the same as that of the Task length treatment. First, participants familiarized themselves with

10.14754/CEU.2022.08

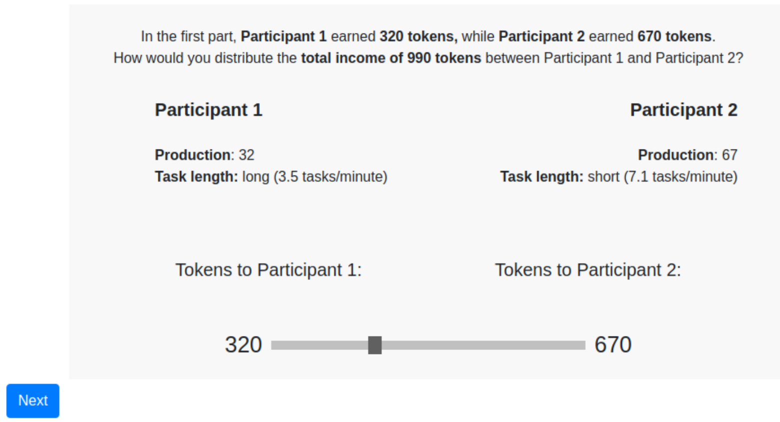


Figure 1.2: Example of a redistribution decision screen

Note: This is an example spectator decision screen in the task length treatment. For stakeholders, their data was presented under the title 'You'. Spectators and stakeholders in the ability treatment saw 'ability group' instead of 'task length' on the screen, with labels 'low', 'medium', and 'high' instead of 'long', 'medium', 'short'.

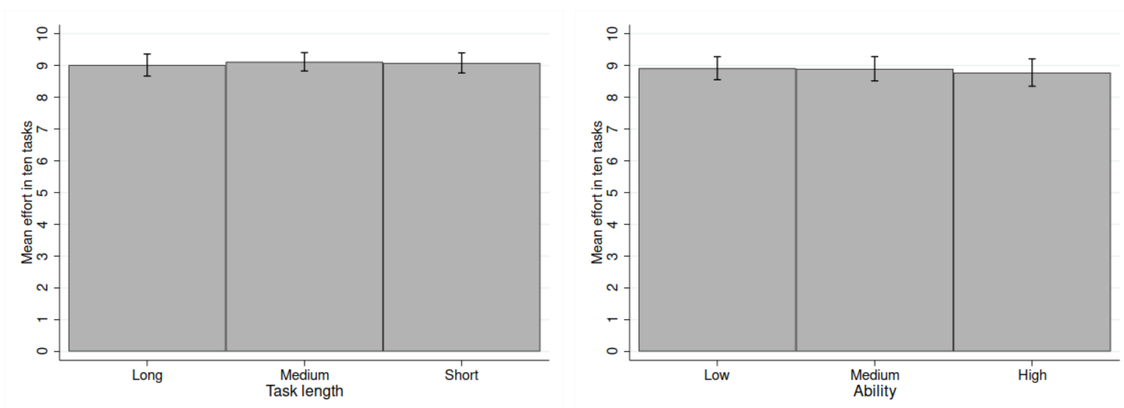
the task, then, after the consent page they did 10 tasks, as fast as they could. Here everyone's tasks were 3-letter tasks, so the variation in difficulties came from own ability in the task. Although this task does not require a specific knowledge or special skills, participants who type faster or can find patterns faster on the screen can perform it better. Based on how fast participants solved the 10 tasks, I categorized them into ability terciles: low, medium, and high. I calculated the mean tasks/minute in each tercile, to have a measure of the relative difficulty of the task across the ability groups, similarly to that in the task length groups. Then, participants performed the task for 15 minutes. Their income from the first part was determined the same way as in the Task length treatment: they received 10 tokens for each correctly solved task.

It is possible that ability in the task correlates with other characteristics, such as age, education level, or employment history. This does not affect my identification if the participants do not consider these characteristics relevant for redistributive decisions. If, for example, they think that older people are poorer, and they are also slower in this task, they might give more money to low-ability subjects not to compensate them for low ability, but because these subjects might be in greater need of the money. In Section 1.3.1 I look at the correlates of the ability measure with the available demographic characteristics. While ability correlates with age and some labor market characteristics, it is unlikely that subjects connect ability with the deservingness of these characteristics. I discuss this in more detail in Section 1.4.

Another concern might be that what I measure as ability here also incorporates effort, as how fast people can perform 10 tasks naturally depends on how hard they

10.14754/CEU.2022.08

try. I tried to minimize the differences in the effort people put in the 10 tasks stage in both treatments by asking them to do them as fast as they can, and by emphasizing that their performance will provide important information about how fast these tasks can be done. There was also a time limit on this part, and participants were informed that they would be excluded from the experiment if they did not finish the 10 tasks within the time limit (see the exact instructions in the Appendix). At the end of the experiment I asked all participants to answer a few questions about how hard they worked on different parts of the experiment, and about their fairness views (see a screenshot of the survey questions in Appendix Figure A.1). Concerning their effort level, they had to rate on a 0 to 10 scale how hard they worked in each stage (10 tasks and production) of the first part. Figure 1.3 shows the self-reported effort levels in the 10 tasks stage in the task length treatment and the ability treatment. The mean effort levels do not differ significantly by either task length or ability group, which suggests that the effect of effort in the ability measure is negligible.



(a) Task length treatment

(b) Ability treatment

Figure 1.3: Mean self-reported effort level by group

Note: The figures show the means of the self-reported effort level of participants when doing 10 tasks as fast as they can, by group (task length or ability). The lines show 95 percent confidence intervals.

1.2.2.2 Redistribution The redistribution stage was also similar to that of the Task length treatment. Except, here participants learned the production levels and the ability groups – low, medium, or high – of the two members of the pair, and the mean tasks/minute people could do in the particular ability group. I decided to present the mean difficulty of the task by ability group to make the ability measure comparable to the difficulty measure in the task length treatment. Here, similarly to the task length treatment, if someone cares about the difficulties in producing when making a redistributive decision, she compares the two participants' production to how difficult the tasks were to them in expectation. The production level of a participant compared to the average tasks/minute in her ability group depends on her effort but also on her ability compared to the average. Appendix Figure A.2 shows the distribution of the individual tasks/minute measure by task length and ability group. The within-group variance of the measure is higher in the task length group, and the highest for people with short tasks, so we have to note that individual ability within group affects the production achieved in the task length treatment more than in the ability treatment.

1.2.3 Participants

I recruited participants currently residing in the United States. They earned about 8.5 USD on average in total. The participants are not representative to the US population, so I present in Table 1.1 their demographic characteristics compared to the US population. The average age in the sample is 35 years, and the median is 33, which is lower than the median age in the U.S, 38.5. There is a larger share of females in my sample than in the population (56 percent vs. 51 percent). There is a higher share of immigrants (80 percent born in the US vs. 85 percent in the population). There are more students among the participants than in the US population (31 percent vs. 10 percent among the 18-year-olds or above). The share of employed is similar to the population (62 vs 61 percent), but the shares of unemployed and out of the labor force are different (16 vs 3 percent and 20 vs 36 percent, respectively).

	Mean	SD	N	Pop. mean
Age	34.93	(13.10)	585	38.50*
Female	0.56	(0.50)	589	0.51
Born in the US	0.80	(0.40)	582	0.85
Currently studying	0.31	(0.46)	484	0.10
Employment status				
Employed (part-time or full-time) or about to start employment	0.62	(0.49)	458	0.61
Unemployed	0.16	(0.37)	458	0.03
Not in labor force	0.20	(0.40)	458	0.36

Table 1.1: Demographic characteristics of participants

Note: Demographic characteristics of the participants. Participants give these data and other background information to Prolific upon registration to the platform. The data presented here were available for the researcher to download. Some participants revoked their consent for the researcher to see the data or the data expired by the time of the experiment, hence the varying number of observations across the rows. Source of the US population statistics: United States Census data, 2019 (https://data.census.gov/cedsci/). The labor market status data from the Census covers people aged 16 or above, while in my sample, the minimum age is 18. *Median age is presented in the US population instead of the mean age.

1.3 Descriptive statistics

1.3.1 Production

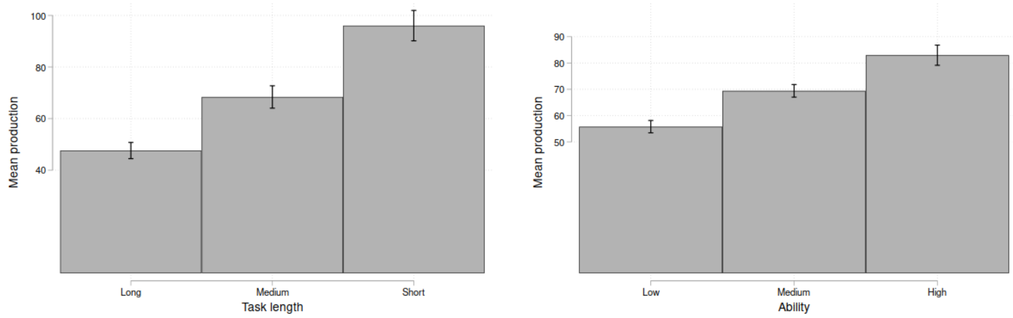
There were 302 participants in the task length treatment, and 292 participants in the ability treatment. According to how fast participants solved the 10 tasks at the beginning of the first part, I computed the mean tasks per minute participants in a group (task length or ability tercile) could do. Table 1.2 shows the group means in both treatments. People with long tasks or low ability are indicated as the Low tasks/min group, with medium tasks length or ability as the Medium tasks/min group, and people with short tasks or high ability are the High tasks/min group. The exogenous task length introduced higher variation in the tasks/minute than people's own abilities. The group means in the Medium tasks/min groups are the same (4.74 tasks) across the treatments, since both groups had 3-letter tasks. Short tasks allowed people to do more tasks per minute than high ability in the 3-letter tasks (6.54 vs 6.00 , $p = 0.000$), and long tasks slightly less than low ability in the 3-letter tasks (3.49 vs 3.72, $p = 0.000$).

Level	Task length experiment	Ability experiment
Low tasks/min	3.489	3.715
Medium tasks/min	4.744	4.737
High tasks/min	6.539	6.001

Table 1.2: Mean tasks/minute by group in each treatment

Note: Each cell shows the number of tasks a group was able to do on average within a minute when asked to do 10 tasks as fast as they can. In the Task length treatment, the groups are determined by their exogenously assigned task length – the Low tasks/min group includes participants with long tasks, the High tasks/min group participants with short tasks. In the Ability treatment, the groups are terciles of the distribution of how fast participants could solve the 10 tasks. The Low tasks/min group includes participants in the lowest tercile, and the High tasks/min group participants in the highest tercile.

Figures 1.4a and 1.4b reveal that the relative difficulty of the task between the groups translate to differences in production, too. The variance of production was also higher in the task length treatment than in the ability treatment. In the task length treatment, the mean production of people with long tasks was 47.6 tasks, with medium tasks 68.4, and with short tasks 96.1. The means in the ability treatment are less spread out, but still significantly different from each other: they are 55.8 tasks for the low ability group, 69.4 tasks for the medium, and 83 for the high ability group. In the survey at the end, participants reported to have worked hard in both treatments and in both stages: on a 0-10 scale, 9.07 (10 tasks) vs 9.29 (production, diff p=0.008) in the task length treatment, and 8.86 (10 tasks) vs 9.01 (production, diff p=0.149) in the ability treatment. The median answer to all four questions was 10. Controlling for the self-reported level of effort, own ability or task length still has a large effect on how much participants could produce, and therefore, in a meritocratic view, how much income they deserve (see Appendix Figures A.3a and A.3b and Appendix Table A.1 for more detail).



(a) Task length treatment (b) Ability treatment

Figure 1.4: Group mean production by treatment

Note: The figures show the mean tasks participants did within the 15 minutes of production in both treatments. The groups within the treatments are the exogenous task length groups in the Task length treatment and the ability terciles in the Ability treatment. The bars show 95 percent confidence intervals.

1.3.1.1 Ability measure The way I measure ability might be correlated with some demographic characteristics. If the subjects think that people with these characteristics are poorer, and in need of more money, they might redistribute more to low-ability subjects because of these beliefs, and not for compensating for low ability. Table 1.3 looks at how the time taken for the 10 tasks in the ability treatment correlates with demographic characteristics. Subjects at the average age (35 years) who are full-time employed take about 2 minutes for the 10 tasks. Each additional year of age adds about half a second to the time taken. Part-time workers and currently unemployed participants also take more time by 18 and 24 seconds, respectively. There is no difference in the time taken by gender, student status, or by whether someone was born in the US.

1.3.2 Redistribution

500 participants completed the second part, which means a 16 percent attrition compared to the first part. Not coming back to the second part did not correlate with either any feature of the experiment or any demographic characteristic (see Appendix Table A.2). Figure 1.5 shows the share of tokens allocated to a randomly chosen participant in the pair by spectators, and the share of tokens allocated to

	Time taken for the ten tasks in seconds
Age	0.535***
	(0.191)
Female	0.0132
	(4.254)
Currently studying	-2.158
	(5.056)
Born in the US	1.689
	(4.877)
Employment status, ref.: Full-time employed	
Due to start a new job within the next month	-6.719
	(11.76)
Not in paid work (e.g. homemaker', 'retired or disabled)	8.031
	(6.639)
Other	7.798
	(8.059)
Part-Time	17.71***
	(5.781)
Unemployed (and job seeking)	24.41***
	(5.727)
Constant	120.8***
	(5.213)
Observations	225

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Time taken for the 10 tasks in the ability treatment

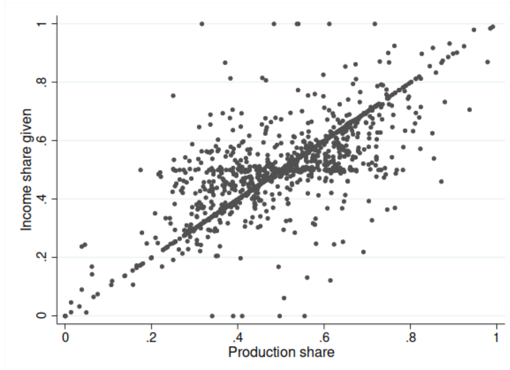
Note: Time is measured in seconds. Age is demeaned.

self by stakeholders, compared to the same participant's production share.[2] In the spectator figures two types of allocations can be clearly distinguished: the egalitarian ones – giving half of the tokens to both participants – and the meritocratic ones – giving a share of tokens equal to the production share. In the stakeholder figures a third type of decision is also salient: giving all income to themselves. A significant portion of decisions is, however, outside these clearly defined shares, which, in the stakeholder figures, might only indicate partly selfish stakeholders giving a little more to themselves than half or than their production share. However, in the spectator figures these allocations suggest that some participants might follow other rules than those of the egalitarian or the meritocratic fairness views. It is also visible that the allocations in the Ability treatment are less spread out and less different from the meritocratic, egalitarian, and fully selfish allocations than in the Task length treatment.

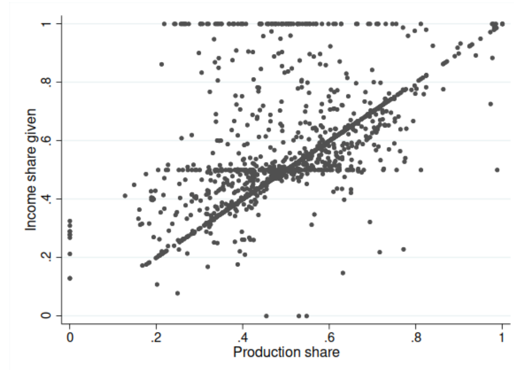
My main question is how relative difficulties within the pair affect the allocation decisions. Figure 1.6 shows the excess share of income allocated to a random participant in the pair on top of her production share in three situations: when she

[2] I chose to define spectators' choices as the share allocated to a random participant in the pair to be able to compare their choices directly to stakeholder choices. Alternatively, I could have used the share given to the participant with lower initial income in the pair (as in e.g., Almås et al., 2020). Since the allocations within a pair are symmetric, we can choose either definition without loss of generality.

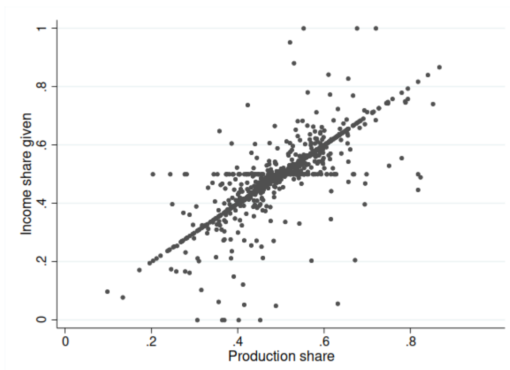
10.14754/CEU.2022.08



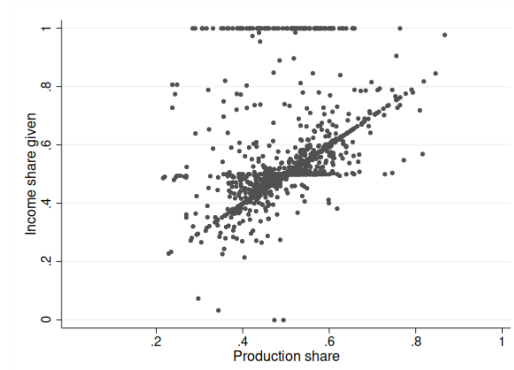
(a) Task length treatment, spectators



(b) Task length treatment, stakeholders



(c) Ability treatment, spectators



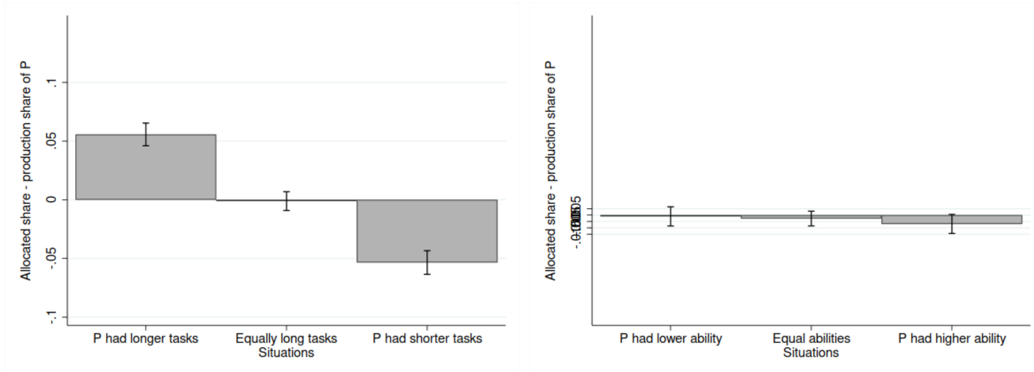
(d) Ability treatment, stakeholders

Figure 1.5: Share allocated to a random participant in the pair (spectators) / to self (stakeholders)

Note: The figures show the share of tokens allocated to a randomly chosen participant in the pair by spectators and to self by stakeholders, plotted against the production share of the same participant. The upper two figures show the spectator and stakeholder decisions in the Task length treatment, and the lower two in the Ability treatment.

had longer tasks/lower ability than her partner, when the task lengths/abilities were equal, and when she had shorter tasks/higher ability. In the task length treatment, as shown in Figure 1.6a, participants who had longer tasks within the pair receive around a 5 percentage points higher income share than their production share, and, symmetrically, those who had shorter tasks receive 5 percentage points less. Participants in pairs with equally long tasks do not receive more or less on average than their production share. In contrast, in the ability treatment, shown in Figure 1.6b, we see no such compensation for differences in ability: participants on average do not get more or less than their production share.

Figure 1.7 shows the analogous comparison of stakeholder decisions: the excess income share given to themselves by stakeholders on top of their production share,



(a) Task length treatment (b) Ability treatment

Figure 1.6: Excess income share allocated by situation (spectator decisions)

Note: The figures show the excess income share given to a randomly chosen participant in the pair on top of her production share by spectators, in three situations. The situations are distinguished by the relative difficulty of the randomly chosen participant.

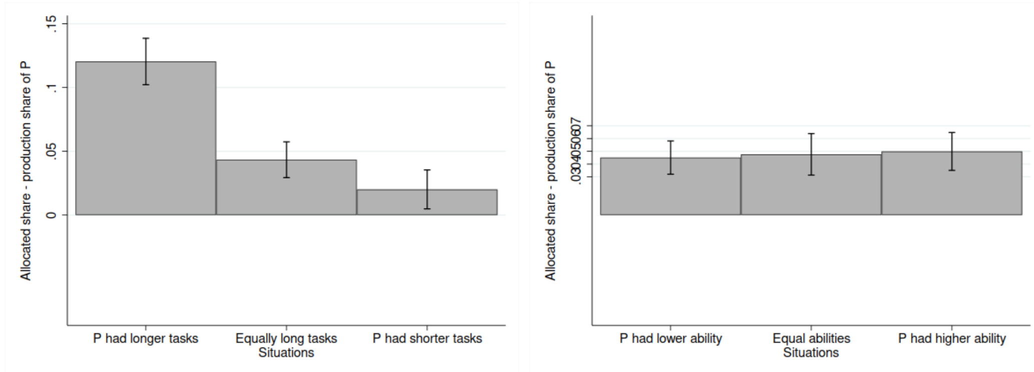
in three situations. The situations are distinguished by the relative difficulty of the decision-maker stakeholder. As we can see in Figure 1.7a, stakeholders who had longer tasks give themselves a 12 percentage points higher income than their production share, while stakeholders who had shorter tasks also give themselves more than their production share, although by much less, only 2 percentage points. Stakeholders in situations where the task lengths were equal gave themselves around 4.5 percentage points higher income than their production share. This suggests that stakeholders with longer tasks might compensate themselves much more than by how much stakeholders with shorter tasks compensate them. In Figure 1.7b we can see that in the ability treatment stakeholders behave similarly in all three situations: they allocate around a 4.5 percentage points higher income to themselves than their production share, no matter the ability differences. Appendix Figures A.4 and A.5 show that the comparisons are similar if the distributive choices are not grouped by situations but by the task length or ability group of the participant receiving the income.In the next section, I explore these relationships in a reduced-form regression framework, to see if the results are robust to different specifications.

1.4 Reduced-form analysis

First, I look at how the excess income share given on top of the production share to a random participant in the pair depends on the relative difficulties within the pair. Let us call the randomly chosen participant Participant 1 (P1), and the other member of the pair Participant 2. Then, for spectators, I run the following regression:

$$s_{1,j,p} = \alpha_0 + \alpha_1 \cdot x_{sh,1,p} + \alpha_2 \cdot (\theta_1 < \theta_2)_p + \alpha_3 \cdot (\theta_1 > \theta_2)_p + \epsilon_{j,p} \tag{1.1}$$

10.14754/CEU.2022.08



(a) Task length treatment

(b) Ability treatment

Figure 1.7: Excess income share allocated by situation (stakeholder decisions)

Note: The figures show the excess income share given by participants to themselves on top of their production share, in three situations. The situations are distinguished by the relative difficulty of the decision-maker stakeholder.

$s_{1,j,p}$ is the share of tokens spectator j allocates to Participant 1 on top of her production share from the joint number of tokens of the pair. $x_{sh,1,p}$ is the production share of Participant 1 in pair p, θ_i is the average number of tasks people in the task length / ability group of Participant i can do within a minute. $\theta_1 < \theta_2$ is therefore the situation where P1 had a harder job than P2, and $\theta_1 > \theta_2$ is when P1 had an easier job. Table 1.4 shows the results of this regression in the task length treatment.

Columns 1-4 of Table 1.4 present relative difficulty in categories. Spectators on average redistribute 5.7 percent of the total income to participants with longer tasks in the pair, and they take 5.5 percent of the total income from participants with shorter tasks in the pair. Column 2 controls for the production share of the participant, too, as from a higher production share – and therefore initial income share – there is less scope for allocating an even higher income. Column 3 adds basic demographic controls: age, gender, whether the spectator was born in the US, and whether she has US nationality.[3] Column 4 adds participant fixed effects, to control for any individual-specific allocation behavior that does not depend on the production share or the relative difficulties. We can see in columns 2 to 4 that the higher the production share of the participant, the less excess income she receives on top of her production share. However, the effects of relative difficulties remain: at a given production share, spectators give a 3.6-4 percentage points higher income if the participant had longer tasks than the other member of the pair, and 3.8-4.1

[3]Student status and employment status were also available in the Prolific database. However, there is a larger share of missing values in these variables, as for some participants these data had expired by the time of the experiment. Appendix Tables A.3-A.6 show the reduced-form results on the sample of participants for whom I have all demographic data.

	Excess income share to random participant in pair							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equally long tasks								
P had longer tasks	0.0574***	0.0397***	0.0400***	0.0366***				
	(0.00781)	(0.00719)	(0.00722)	(0.00773)				
P had shorter tasks	-0.0555***	-0.0381***	-0.0377***	-0.0412***				
	(0.00733)	(0.00697)	(0.00699)	(0.00812)				
Production share		-0.154***	-0.154***	-0.151***		-0.144***	-0.145***	-0.142***
		(0.0303)	(0.0305)	(0.0326)		(0.0313)	(0.0316)	(0.0341)
Relative difficulty					0.0267***	0.0189***	0.0188***	0.0187***
					(0.00293)	(0.00277)	(0.00277)	(0.00297)
Constant	0.00473	0.0828***	0.0851***	0.0774***	0.00692	0.0795***	0.0822***	0.0719***
	(0.00607)	(0.0146)	(0.0160)	(0.0169)	(0.00580)	(0.0149)	(0.0163)	(0.0170)
Observations	1210	1210	1210	1210	1210	1210	1210	1210
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Spectator decisions in the task length treatment

Note: The outcome variable is the excess income share given to a randomly chosen participant (Participant 1) in a pair on top of her production share by spectators in the task length treatment. Column 1 controls only for the situation of the participant receiving the income – whether she had longer tasks, equally long tasks, or shorter tasks than the other member of the pair (same as Figure A.4a). Columns 2-4 also control for the production share of Participant 1. Column 3 includes demographic controls: age, gender, whether the spectator was born in the US, and whether she has US nationality. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the task length groups of P2 and P1 can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

percentage points if she had shorter tasks. The constant is significant and positive, meaning that on average, participants receive some excess income from spectators, unconditional on their production share. Columns 5-8 present relative difficulty as a continuous measure, namely, by how much more tasks Participant 2's task length group could do than Participant 1's group on average. Each one task per minute disadvantage gives a 2.7 percentage points higher income to participants on average, and a 1.8 percentage points higher income controlling for their production share.

Table 1.5 shows the result of the same regression in the ability treatment. Relative difficulties do not matter in the decisions of spectators in the ability treatment, either on average, or when controlling for production share. In general, spectators seem to condition their allocation decisions much more strongly on production share and less on any other factors. Participants do not receive significantly higher income on average than their production share, as shown by the smaller and insignificant constant in the regressions. The higher the production share, the less excess income participants receive from spectators, but this relationship is also much weaker than in the task length treatment. Referring back to the discussion in Section **??**, while the ability measure correlates with age and labor market status, subjects do not compensate for lower ability at all in their redistributive decisions. Since there is no compensation whatsoever, it is unlikely that subjects thought about ability from the perspective of fairness related to characteristics it is correlated with.

	Excess income share to random participant in pair							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equal ability								
P had lower ability	0.00137	-0.00516	-0.00496	-0.00448				
	(0.00719)	(0.00714)	(0.00716)	(0.00787)				
P had higher ability	-0.00477	0.00185	0.00202	0.00246				
	(0.00779)	(0.00651)	(0.00662)	(0.00781)				
Production share		-0.103**	-0.102**	-0.0938*		-0.105**	-0.103**	-0.0966*
		(0.0476)	(0.0476)	(0.0516)		(0.0470)	(0.0470)	(0.0514)
Relative difficulty					0.00191	-0.00230	-0.00225	-0.00238
					(0.00291)	(0.00226)	(0.00224)	(0.00248)
Constant	-0.00762	0.0448*	0.0243	0.0437	-0.00874**	0.0447*	0.0241	0.0445*
	(0.00606)	(0.0246)	(0.0248)	(0.0273)	(0.00439)	(0.0231)	(0.0233)	(0.0256)
Observations	1170	1170	1170	1170	1170	1170	1170	1170
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Spectator decisions in the ability treatment

Note: The outcome variable is the excess income share given to a randomly chosen participant (Participant 1) in a pair on top of her production share by spectators in the abiltiy treatment. Column 1 controls only for the situation of the participant receiving the income – whether she had lower ability, equal abiltiy, or higher ability than the other member of the pair (same as Figure A.4b). Columns 2-4 also control for the production share of Participant 1. Column 3 includes demographic controls: age, gender, whether the spectator was born in the US, and whether she has US nationality. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the ability groups of P2 and P1 can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

To look at the behavior of the stakeholders, I run the following regression:

$$s_{own,j,p} = \alpha_0 + \alpha_1 \cdot x_{sh,own,j} + \alpha_2 \cdot (\theta_{own} < \theta_{other})_p + \alpha_3 \cdot (\theta_{own} > \theta_{other})_p + \epsilon_{j,p} \quad (1.2)$$

Here, $s_{own,j,p}$ is the share of tokens stakeholder j gives to herself on top of her production share in the decision characterized by pair p. $x_{sh,own,j}$ is the production share of stakeholder j in pair p. The relative difficulties are expressed from the perspective of the decision-maker stakeholder: $\theta_{own} < \theta_{other}$ if she had a harder job, and $\theta_{own} > \theta_{other}$ if her partner had a harder job. Table 1.6 shows the results of this regression in the task length treatment.

The picture here is not as clear as for the spectators, because in each decision, selfishness concerns are involved besides fairness considerations. Stakeholders give themselves an 8.2 percentage points higher income if they had longer tasks in the pair, and a 2 percentage points lower income if they had shorter tasks, than when the task lengths were equal. However, when we control for production share, demographics, and participant fixed effects – which partly take care of the effect of selfishness on the allocation decision –, only the self-compensation of long-task stakeholders remain: they give themselves 2.5-4.7 percentage points more of the total tokens. Stakeholders

	Excess income share to self							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equally long tasks								
P had longer tasks	0.0828***	0.0476**	0.0454**	0.0252***				
	(0.0206)	(0.0201)	(0.0194)	(0.00824)				
P had shorter tasks	-0.0214*	0.0154	0.0138	0.000523				
	(0.0129)	(0.0161)	(0.0160)	(0.00861)				
Production share		-0.336***	-0.349***	-0.307***		-0.342***	-0.351***	-0.290***
		(0.0674)	(0.0667)	(0.0514)		(0.0699)	(0.0694)	(0.0534)
Relative difficulty					0.0238***	0.00586	0.00650	0.00767***
					(0.00598)	(0.00700)	(0.00715)	(0.00259)
Constant	0.0300	0.196***	0.241***	0.210***	0.0531**	0.222***	0.262***	0.211***
	(0.0213)	(0.0371)	(0.0702)	(0.0251)	(0.0224)	(0.0405)	(0.0709)	(0.0269)
Observations	1250	1250	1250	1250	1250	1250	1250	1250
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Stakeholder decisions in the task length treatment

Note: The outcome variable is the excess income share given to self in a pair on top of her production share by stakeholders in the task length treatment. Column 1 controls only for the situation of the decision-maker stakeholder – whether she had longer tasks, equally long tasks, or shorter tasks than the other member of the pair (same as Figure A.5a). Columns 2-4 also control for the production share of the stakeholder. Column 3 includes demographic controls: age, gender, whether she was born in the US, and whether she has US nationality. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the task length groups of the other pair member and the stakeholder can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

also give themselves a large excess income unconditional on their production share, as the large and significant constant shows. When we look at relative difficulty as a continuous measure in columns 5-8, the effect of relative difficulty on the excess allocated income share is quite unstable, most likely because of selfishness concerns playing a role in the decisions. Stakeholders on average give themselves 2.4 percentage points more of the income with each one task disadvantage, but when we include participant fixed effects, this reduces to less than 1 percentage point.

In the ability treatment, as shown in Table 1.7, stakeholders give themselves 4 percentage points more than their production share on average, but this is not different by the relative difficulty of the decision-maker stakeholder. When controlling for their production share and demographic controls, higher-ability stakeholders seem to give themselves more than in equal-ability and lower-ability situations, but these results are not robust to including participant fixed effects, so controlling for individual selfishness. Looking at relative difficulty as a continuous measure in Columns 5-8 gives similar results. All in all, stakeholders in the ability treatment do not seem to compensate themselves or each other for low ability, either. In the next section, I introduce the model of fairness preferences by Cappelen et al. (2010), that allows us to identify fairness preference types in the sample and look at the results in a structural framework.

	Excess income share to self							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equal ability								
P had lower ability	-0.00430	-0.0323	-0.0279	0.00563				
	(0.0184)	(0.0233)	(0.0214)	(0.00736)				
P had higher ability	0.00183	0.0326**	0.0298**	0.0120				
	(0.0155)	(0.0153)	(0.0143)	(0.00964)				
Production share		-0.504***	-0.495***	-0.195***		-0.526***	-0.517***	-0.189***
		(0.173)	(0.153)	(0.0508)		(0.187)	(0.165)	(0.0506)
Relative difficulty					-0.00133	-0.0215*	-0.0195*	-0.00110
					(0.00834)	(0.0120)	(0.0107)	(0.00357)
Constant	0.0424*	0.299***	0.411***	0.141***	0.0416*	0.311***	0.423***	0.144***
	(0.0253)	(0.0955)	(0.102)	(0.0265)	(0.0249)	(0.0978)	(0.104)	(0.0253)
Observations	1210	1210	1210	1210	1210	1210	1210	1210
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Stakeholder decisions in the task length treatment

Note: The outcome variable is the excess income share given to self in a pair on top of her production share by stakeholders in the ability treatment. Column 1 controls only for the situation of the decision-maker stakeholder – whether she had lower, equal, or higher ability than the other member of the pair (same as Figure A.5b). Columns 2-4 also control for the production share of the stakeholder. Column 3 includes demographic controls: age, gender, whether she was born in the US, and whether she has US nationality. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the ability groups of the other pair member and the stakeholder can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

1.5 Structural analysis

1.5.1 Descriptive model

I use the model of fairness preferences by Cappelen et al. (2010), applied to my setting, as the base for the structural analysis. According to the model, a spectator with fairness preference type k finds the following allocation of joint income Y fair:

$$t_1^k(\mathbf{x},\, \theta) = \frac{f^k(x_1, \theta_1)}{f^k(x_1, \theta_1) + f^k(x_2, \theta_2)} Y(\mathbf{x}(\theta)) \tag{1.3}$$

Here $t_1^k(\mathbf{x},\, \theta)$ is the tokens a spectator with fairness preference type k finds fair to give to Participant 1 in the pair – in my analysis a randomly selected member of the pair. $f^k(x_i, \theta_i)$ is a function that shows how a spectator with fairness preference k values the contribution of pair member i. x_i is the production level of pair member i, and θ_i is the average number of tasks the ability or task length group participant i belongs to can perform within a minute. Let us denote the *share of tokens* a spectator with fairness preference type k finds fair to allocate to P1 in the pair with $s_1^k(\mathbf{x},\ \theta)$:

$$s_1^k(\mathbf{x},\ \theta) = \frac{f^k(x_1, \theta_1)}{f^k(x_1, \theta_1) + f^k(x_2, \theta_2)} \tag{1.4}$$

The most common types of fairness preferences that are distinguished in the literature and can be found in my setting are the following:

- *Meritocratic:* accepts inequality based on merit

$$s_1^k(\mathbf{x},\ \theta) = \frac{x_1}{x_1 + x_2} \tag{1.5}$$

- *Egalitarian:* does not accept any inequality

$$s_1^k(\mathbf{x},\ \theta) = \frac{1}{2} \tag{1.6}$$

Participants with meritocratic fairness preferences find inequalities based on merit fair, so they allocate the tokens proportionally to production. The egalitarian type does not accept any inequality, so she will always give half of the tokens to both members of the pair unconditionally. There is a third common type, the *libertarian* type, that finds all kinds of inequalities fair. In my design, since there is no pure luck component in the income, this type cannot be distinguished from the meritocratic type. I made this simplification to better be able to look at deviations from the meritocratic allocations. Because I want to test if there are people who follow a different rule by compensating for difficulties, I introduce a fourth type:

- *Meritocratic who compensates for difficulties:*

$$s_1^k(\mathbf{x},\ \theta) = \frac{x_1/\theta_1}{x_1/\theta_1 + x_2/\theta_2} \tag{1.7}$$

This type wants to reward production, but she weighs the production of the members of the pair with how easy or difficult it was for them to produce. A good proxy of this type of decision would be that the weights are the inverse of the number of tokens the task length or ability group of person i can do within a minute (θ_i). I do not assume that participants who want to compensate have this exact rule in mind, but it is a good starting point for separating compensation from purely meritocratic decisions. In the ability treatment this person compensates for low ability in the task, so she rewards pure effort. When facing a decision over the income of two participants from the same group, however, she behaves like a pure meritocrat: she rewards higher performance, which partly includes higher ability within group, as well. In the task length treatment a compensating meritocrat compensates for external difficulties. She rewards performance but she gives more tokens to the participant who had harder tasks, so she also wants to reward effort. When comparing two participants from the same task length group, she rewards performance, which, as in the ability treatment, depends on effort but also on ability *within* a task length group. However, as we saw in Appendix Figure A.2, the variance of the tasks/minute within group is higher in the task length treatment, and it is quite small in the ability treatment, we can say that a compensating meritocrat rewards effort and ability but compensates for external difficulties in the task length treatment, and she rewards effort and compensates for ability differences in the ability treatment.

We assume that both spectators and stakeholders fall into one of these types, but their optimization problems in a redistributive decision are different. The utility maximization of spectators is the following (from Cappelen et al., 2010, 2020):

$$U_j(\mathbf{x},\ \theta, k_j) = -(t_{1,j} - t_1^{fair,k_j})^2 \tag{1.8}$$

Here $t_{1,j}$ is the decision variable of spectator j – the number of tokens allocated to Participant 1 in the pair, while t_1^{fair,t_i} is what she finds fair to allocate given her fairness preferences. Since spectators do not have any monetary gain from the decision, they simply choose the allocation that aligns with their fairness preferences. The optimal spectator decision is therefore:

$$t^*_{1,j} = t_1^{fair,k_j} \tag{1.9}$$

Stakeholders, on the other hand, consider both their own monetary gain and the fairness of the allocation. Their problem can be written as follows:

$$U_i(\mathbf{x},\ \theta, k_j, \beta_j) = t_{own} - \beta_j \frac{(t_{own,j} - t_{own}^{fair,k_j})^2}{2Y}, \tag{1.10}$$

Here, $t_{own,j}$ is the number of tokens stakeholder j gives to herself in the decision, while t_{own}^{fair,k_j} is the tokens she finds fair to give if she has a fairness preference type k. β_j is the weight she puts on fairness relative to her own income. Therefore, the optimal stakeholder decision is:

$$t^*_{own,j} = t_{own}^{fair,k_j} + \frac{Y}{\beta_j} \tag{1.11}$$

Stakeholders give themselves the amount of tokens they find fair and some extra tokens depending on how selfish they are.

1.5.2 Structural analysis

Based on the model described above, I use a simple, individual-level categorization into one of the fairness types. Since everyone made 10 decisions, we can observe one person's choices in many different situations. In every decision, we know what a person with meritocratic preferences, with egalitarian, and with compensating meritocratic preferences would choose. Therefore, for each individual, I calculate the sum of squared deviations in each decision from what someone with a particular fairness type would have chosen. I categorize each participant into a type that minimizes the sum of squared deviations:

$$\text{type}_j = \arg\min_k (t_{1,j} - t_{1,k})^2 + (t_{2,j} - t_{2,k})^2 + ... + (t_{10,j} - t_{10,k})^2, \tag{1.12}$$

where $k =$ Meritocratic, Egalitarian, or Compensating meritocrat. Then I compute the share of each type separately for the task length and the ability treatment. To obtain standard errors for the shares, I use the bootstrap method: I draw 10 random decisions with replacement from the 10 decisions of each individual 1000 times, and calculate the shares of the types using the individual categorizations with the 10 random draws. The above categorization works well for spectators, but for stakeholders, as we saw in the model, selfishness concerns play a role, as well. For now, I use the same categorization for stakeholders as for spectators, but there are some stakeholders who in most decisions allocate all tokens for themselves, so we cannot infer their fairness type. To handle this, I add a fourth category for stakeholders: a fully selfish type, whose decisions are closest to always taking all of the tokens.

Figure 1.8 show the calculated shares for spectators in both treatments. In the ability treatment, 83 percent of spectators are classified as meritocrats, 14 percent as egalitarians, and virtually zero as compensators, just as the reduced-form results suggested. In the task length treatment, on the other hand, only 53 percent are meritocrats, 17 percent are egalitarians, and 29 percent are compensators. The picture is similar when we look at stakeholders in Figure 1.9. In the ability treatment, 69 percent of stakeholders are meritocrats, 21 percent are egalitarians, virtually zero are compensators, and 6 percent are fully selfish. In the task length treatment, 51 percent are meritocrats, 20 percent are egalitarians, 19 percent are compensators, and 9 percent are selfish.

In the reduced-form results it seemed that stakeholders with long tasks compensated themselves more for their difficulties than by how much stakeholders with short tasks compensated them. To test this in the structural framework, I computed the shares of fairness types in each task length group. Figure 1.10 shows that indeed, stakeholders who got long tasks are around 25 percentage points less likely to make meritocratic decisions, and they are more likely to make all other types of decisions. 24 percent in the long-task group make compensating decisions, while only 18 percent in the medium-, and 15 in the short-task group are classified as compensators. Long-task stakeholders are also more likely to make egalitarian and fully selfish decisions than medium- and short-task participants. If this is the case then there is a self-serving choice of fairness norms that is not explained by the model described in Section 1.5.1. Figure 1.11 shows the calculated shares among stakeholders in the

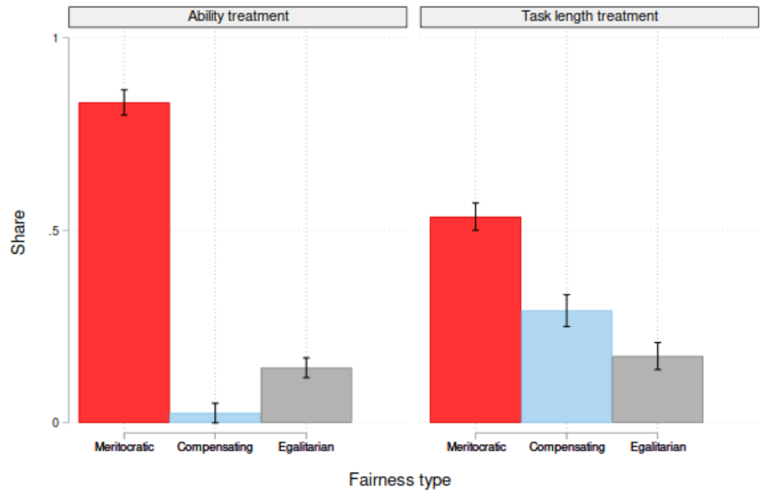


Figure 1.8: Shares of fairness types among spectators

Note: Shares of fairness types among spectators in the task length treatment and the ability treatment. The lines show 95 percent confidence intervals using bootstrapped standard errors.

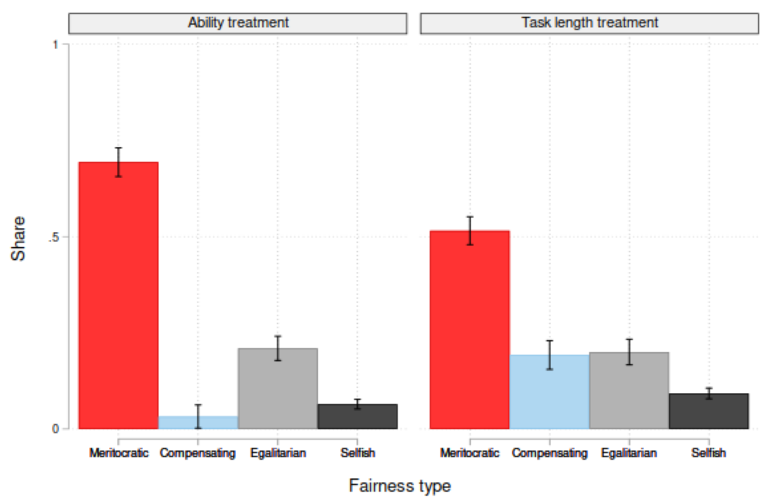


Figure 1.9: Shares of fairness types among stakeholders

Note: Shares of fairness types and the fully selfish type among stakeholders in the task length treatment and the ability treatment. The lines show 95 percent confidence intervals using bootstrapped standard errors.

ability treatment by ability group. Here, all ability groups are equally likely to make meritocratic, compensating, and egalitarian decisions. The only difference between the groups is that high-ability stakeholders seem to make more fully selfish decisions than the other two groups.

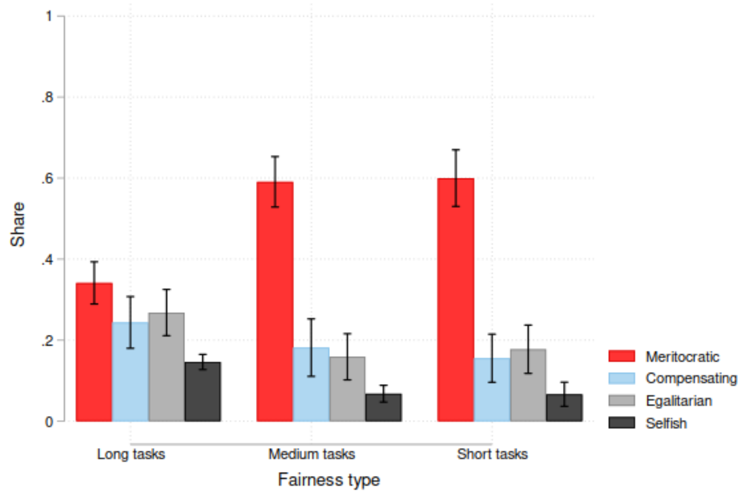


Figure 1.10: Shares of fairness types among stakeholders by task length

Note: Shares of fairness types among stakeholders in the task length treatment by task length. The lines show 95 percent confidence intervals using bootstrapped standard errors.

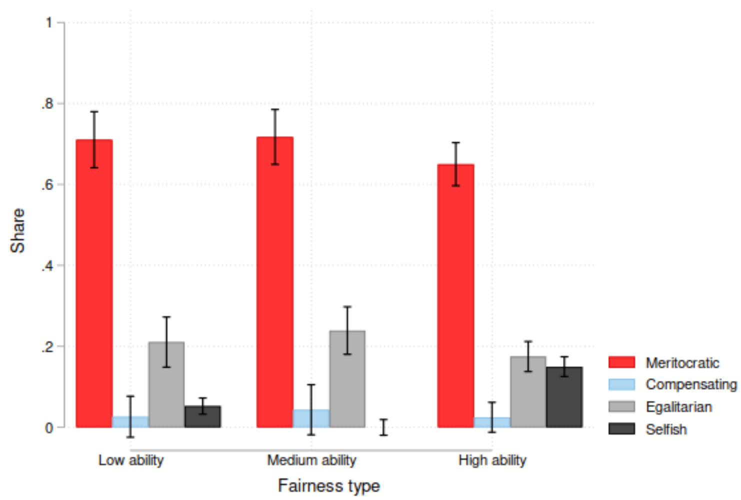


Figure 1.11: Shares of fairness types among stakeholders by ability

Note: Shares of fairness types among stakeholders in the ability treatment by ability group. The lines show 95 percent confidence intervals using bootstrapped standard errors.

Finally, I check if receiving easier tasks simply made participants think in more meritocratic terms, while receiving harder tasks reminded them about external determinants of what one can achieve. Figure 1.12 shows the shares of different fairness types among spectators, separately by the spectator's task length. Surprisingly, the picture here is the exact opposite of that of the stakeholders: spectators who themselves had longer tasks are more meritocratic and less compensating, while spectators with medium and short tasks make less meritocratic and more compensating decisions. Own external circumstances therefore seem to have an effect on what allocations one finds fair, however, the effect is very different depending on whether one's own income is at stake, too. Figure 1.13 shows the shares of fairness types among spectators in the ability treatment by own ability. The shares among ability spectators are exactly the same across the ability groups. Therefore, receiving information about own ability did not change spectators' fairness views.

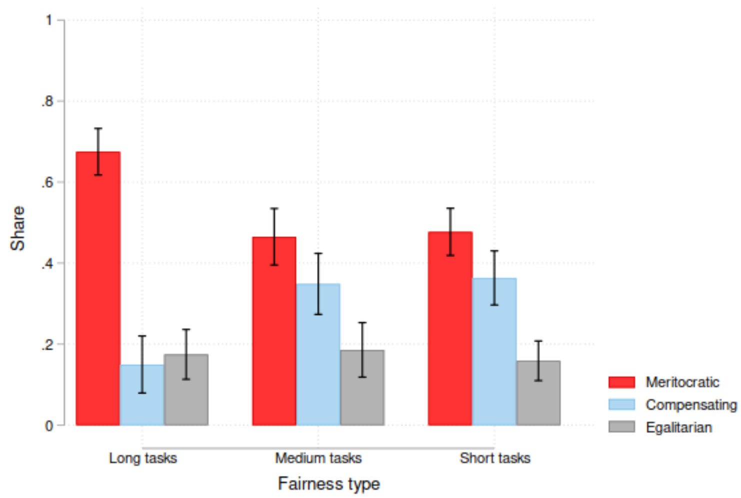


Figure 1.12: Shares of fairness types among spectators by own task length

Note: Shares of fairness types among spectators in the task length treatment by own task length. The lines show 95 percent confidence intervals using bootstrapped standard errors.

1.6 Conclusion

I looked at whether people compensate for differences in how difficult the task was when rewarding production. I studied two different sources of difficulties – an exogenous task length and individual ability in the task. Besides the level of effort participants made in the production, both types of difficulties strongly affected how many tasks participants could produce. I found that participants, both as spectators and as stakeholders, compensated the participant who had a harder job if the hardships were exogenous, but simply rewarded production and ignored the relative difficulties when the hardships arised from differences in individual ability in the task. Based on these results, although one cannot necessarily be made fully responsible for her ability in something – especially in this experiment, where participants were not aware of the task beforehand –, people seem to find income inequalities

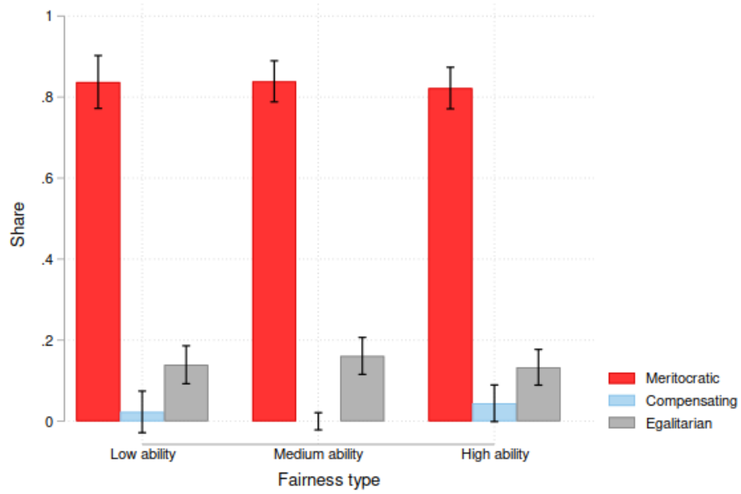


Figure 1.13: Shares of fairness types among spectators by own ability

Note: Shares of fairness types among spectators in the ability treatment by own ability group. The lines show 95 percent confidence intervals using bootstrapped standard errors.

caused by ability differences fair. They seem to draw the line between fair and unfair inequalities along external and internal factors in someone's performance, and not between exogenous and endogenous ones.

I also found evidence on self-serving choice of fairness principles. In the task length treatment, stakeholders with long tasks were more likely to take all of the tokens to themselves, to compensate themselves for hardships, or to divide the tokens equally, than stakeholders with medium or short tasks. They were also much less likely to make purely meritocratic decisions than the other two groups. This suggests that being able to produce more partly because of luck when being assigned shorter tasks made participants believe that they deserve more for their higher production, even though they learned that their partners had a harder job. On the other hand, being assigned long tasks made participants more selfish or choose fairness principles that allowed them higher earnings than what they received for their performance.

10.14754/CEU.2022.08

2 Chapter 2: Excuse-driven Present Bias

Joint work with Marc Kaufmann

2.1 Introduction

There are many reasons why people rightly postpone unpleasant activities. Take a software engineer who has a list of bugs to fix and features to implement. She may wait to hear back from a client to see if the bug persists, check if her colleague still cares about the feature, or decide to leave work for tomorrow when her schedule is free. When applied coherently, these reasons apply with equal strength no matter whether they justify putting off *immediate* work or *future* work. We explore experimentally whether people instead apply such reasons *asymmetrically*, as if these reasons provide stronger justifications when they allow putting off immediate rather than future work. Such asymmetric use of reasons is what we mean by excuse-making. Present bias, with its time-inconsistent nature, is a prime candidate for excuse-making, since excuses may allow people to leave free rein to their impulsive behavior.[4]

This description highlights the main challenge: for it to be convincing, an excuse must be able to masquerade as a reason that is perceived to be genuine; but then, it will influence behavior even when not used as an excuse. Therefore when we compare the willingness to work for people with and without an excuse, we should find that the former work less, because they have a reason to work less. The same reason should lead to the same change in willingness to work. In two experiments we look at whether instead willingness to work is different when there is a behavior – avoiding present work – for which people can use the reason as an excuse. We provide two types of alternatives for work: *no-excuse* and *excuse* alternatives. We describe our within-subjects design to identify excuse-making based on change in willingness to work in Section 2.2. We then present the the first experiment and the implementation on Amazon Mechanical Turk (MTurk) in more detail in Section 2.3.

In the first experiment, the no-excuse alternative required subjects to work harder in the future for sure, while the excuse alternative provided a 10% chance of the work remaining easy. We took the idea of using a kind of risk as a possible excuse from Exley (2016), who finds that people use risk in the donation reaching the charity as an excuse to donate less. Another source of choosing this excuse was intuition: people often postpone work because deep down they hope that later it will get easier or that they will not have to do it at all in the end. We found that the willingness to

[4] Another reason why we refer to present bias is that we focus on short-term time preferences with future decisions at most several days in the future. As others have noted, it is calibrationally implausible to have measurable discount factors at such short time frames that are due to exponential discounting. For example, a weekly discount of 0.99) (which is virtually indistinguishable from 1.00) leads to a yearly discount factor of $0.99^{52} \approx 0.59$). Thus our claim is that whatever short-term impatience we find is likely to be present bias or some other present-focused behavior, rather than impatience of the exponential discounting variety.

work hard today dropped by \$0.11 ($N = 147$, $p = 0.011$) more than the willingness to work hard at a future date, even though the reason to choose the future option was the same in the two cases: we introduced a chance that work will remain easy in the future. We interpret this phenomenon as excuse-driven present bias. In Section 2.4 we discuss the results in more detail.

Section 2.5 discusses the drawbacks of the implementation on MTurk and describes the second experiment which we ran online with students in the LISER-LAB in Luxembourg. Our aim was to strengthen the first results and our identification on three fronts. The first issue was to complement risk as an excuse with other, cleaner excuses, to rule out all state-dependent preferences with risk and time interactions. The second issue is that our MTurk subjects report working roughly 20 hours per week on MTurk. Suppose, to make the point most clearly, that subjects find our task exactly as unpleasant as other MTurk tasks, and can earn a fixed hourly wage from these other tasks. When deciding how much to work today rather than two days from now, these subjects would base their decisions on which day offers a higher payment per task. Their choices would thus be driven by maximizing earnings, and not involve any time preferences. Moving to a student population would alleviate this concern as students are more likely to substitute our task for unpaid leisure or studying, which are less direct substitutes (and less likely to be equally unpleasant). Third, while in the first experiment we could document excuse-making in short-term preferences, we wanted to be able to separate the effect of the excuse on short-term and long-term preferences, so on present bias and patience. Therefore in the second experiment we tested excuse-making in choices made over work at future dates.

In the second experiment, we modified the risk excuse slightly: here the no-excuse alternative meant *extra* work in the future for sure on top of some required work, while the excuse alternative involved a 10% chance of no extra work. We found non-significant results of similar magnitude as in the first experiment: the willingness to work today dropped by €0.06 ($N = 75$, $p = 0.36$) more than the willingness to work in the future when moving from the no-excuse to the excuse alternative. Besides risk, we tested another excuse: a different type of task. Here, the no-excuse alternative was to do the same task in the future people were already working on. The excuse alternative, however, involved a different kind of task in the future. The idea was that people might rationalize postponing work to the future by just choosing to do the other task, therefore, hiding their *implicit preference*, present bias, behind a choice between tasks (similarly as in the framework described in Cunningham and de Quidt (2022). This excuse did not seem to work: the willingness to work today was not significantly different from the willingness to work in the future when moving from the no-excuse to the excuse alternative (the sign was even negative, €-0.032, $N = 75$, $p = 0.64$). We found non-significant results for the choices over work at future dates, as well.

Section 2.6 presents the results from the second experiment in more detail and discusses why we might have failed to replicate the MTurk results. We suspect that one reason was the difference in the introduction of risk: in the first experiment, to minimize MTurk workers substituting our tasks with other MTurk tasks, we chose to have the same amount of work in all choices. Some tasks were, however, *hard* while others *easy*, and participants decided about the number of hard tasks in their choices. When moving to the lab, we chose to have a certain amount of required tasks in each session, and the choices were about *extra* tasks on top of the required amount instead of hard vs. easy tasks. Part of our failure to replicate the results might have been that students still have substituted our tasks with other activities. The other part was probably lack of power to detect an effect size of similar magnitude to that in the MTurk study. Our lab study lasted four weeks and we had significant attrition, potentially leading to low power. In Section 2.6 we also describe how we want to test if choosing extra tasks or hard vs. easy tasks really makes a difference, and to corroborate our findings from the first experiment. Finally, in Section 3.7 we discuss our results, other possible mechanisms to explain them and how we rule these out, and the potential of excuse-making in present bias.

If people excuse their present bias, this suggests strongly that present bias – or present-oriented behavior – is context-sensitive: the same person may behave in more or less present-biased ways in superficially similar situations. For instance, the more dimensions there are to a choice, the more present-biased a person may act. In addition, having excuses for present bias may be what keeps people from learning that they are present-biased. This in turn could ensure that they remain *naive* about their own present bias, which is in line with several studies that find that most subjects are predominantly naive (Augenblick and Rabin, 2019; Fedyk, 2021). At the same time, it leaves room for learning in situations with clear feedback, such as in Le Yaouanq and Schwardmann (2019).[5] Most situations provide more excuses than our experimental setup: life offers more important and urgent tasks to do, colleagues more requests for help, and Netflix more movies to chill than any lab study can ever hope for. If excuses increase people's present bias, daily life where excuses abound will exacerbate this beyond what we find in the lab.

2.2 Identification of excuse-driven present bias

As mentioned in the introduction, we consider the degree to which the same reason is applied asymmetrically as excuse-making. The challenge is in ensuring that we keep the reason constant, while bringing out the asymmetry. Let us observe that people only make excuses if (i) there is a motivation – conscious or subconscious – to distort the choice; and (ii) there is a potential excuse available. Thus when there is no motivation to distort or no possibility to distort, we will not see excuse-making. The basic idea is therefore to offer people 4 choices, for each combination of motivation to distort ('yes' and 'no') and possibility of distorting ('yes' and 'no').

[5] O'Donoghue and Rabin (1999) highlight how important it is whether people are aware of their present bias (sophistication) or unaware of it (naivete).

The hypothesis we are testing is whether present bias is something for which people have a motivation to distort their choices. We have 2 choices involving tedious tasks today as the high-motivation case, and 2 choices involving tedious tasks in the future as the low-motivation case. For now we assume that we have two alternatives, a *no-excuse alternative* and an *excuse alternative* that allow for small and large choice distortions respectively.

Concretely, we want to find the following indifference points X, X', Y, and Y':

No excuse:

$$\begin{aligned} \text{work in the future} + \$X &\sim \text{no-excuse alternative} \\ \text{work in the present} + \$X' &\sim \text{no-excuse alternative} \end{aligned}$$

Excuse:

$$\begin{aligned} \text{work in the future} + \$Y &\sim \text{excuse alternative} \\ \text{work in the present} + \$Y' &\sim \text{excuse alternative} \end{aligned}$$

Then $\Delta_C = X' - X$ is the difference between willingness to work in the future and in the present when there are no possibilities for making excuses (control), while $\Delta_T = Y' - Y$ is the same difference when there are possibilities for making excuses (treatment). While this identification strategy is applicable in the context of any excuse, we first use the potential of avoiding work altogether. Hence, the no-excuse condition involves having to do extra tasks in the future for sure; and the excuse condition having a 10% chance of not needing to do the extra future tasks. This is based on Exley (2016), who finds that people distort risky choices in charitable giving in a way that is beneficial to them.

The main concern of the design is to take care of the difference in utility between the no-excuse and the excuse alternatives, since a drop in the probability of having to do work should lead to a change in willingness to work. To highlight how the four choices in our design allow us to take care of this concern, let us rewrite the indifference conditions in utility terms. X, X', Y, and Y' are the monetary amounts that make people indifferent between the right- and left-hand side options in each decision. For now, let us assume linear utility in money.[6] Let x_1 be the amount of tasks people have to do in the left-hand side options, and x_2 the amount of tasks in the right-hand side (no-excuse or excuse) options. Denote by $d(x)$ the disutility of doing x tasks, by β the present bias parameter in the no-excuse condition, and by β_E the present bias parameter in the excuse condition. We do not think that people's preferences literally change. Rather we think of this as a reduced-form way to capture the idea that people behave as though they were more present biased in the excuse setting.

[6] Later we discuss how the specific choices participants make take care of a possible concave utility of money.

No excuse (control):

$$\begin{aligned} X &= \beta\, d(x_1) - \beta\, d(x_2) \\ X' &= d(x_1) - \beta\, d(x_2) \end{aligned}$$

Excuse (treatment):

$$\begin{aligned} Y &= \beta\, d(x_1) - 0.90\beta\, d(x_2) \\ Y' &= d(x_1) - 0.90\beta_E\, d(x_2), \end{aligned}$$

We assume for simplicity that people have expected utility preferences, although this is not necessary. Taking the difference between willingness to work in the future over in the present in the two conditions, we get the following:

$$\begin{aligned} \Delta_C &= d(x_1)(1-\beta) \\ \Delta_T &= d(x_1)(1-\beta) + 0.9d(x_2)(\beta - \beta_E) \end{aligned}$$

We finally obtain the difference of these differences,

$$\Delta\Delta = \Delta_T - \Delta_C = 0.9d(x_2)(\beta - \beta_E) \tag{2.1}$$

Our null hypothesis, meaning the absence of excuse-making, is that $\Delta\Delta = 0$. This is the case if $\beta_E = \beta$. This should be true for any risk preferences that are *state-independent*, since the first differences wash out any state-independent effect. The alternative hypothesis is that $\beta_E < \beta$, and $\Delta\Delta > 0$, so there is excuse-driven present bias. Even when we have *state-dependent* preferences, such as when the foregone option affects the choice, we can only get non-zero effects if this state-dependent effect is stronger when the alternative is work today rather than work in the future.

2.3 First experiment

2.3.1 Technical details

We implemented our first experiment in oTree (Chen et al., 2016b) and ran it on the Amazon Mechanical Turk (MTurk) platform. All details were pre-registered in the AEA RCT Registry (Drucker and Kaufmann, 2019) We ran the first pilot of the actual experiment in August 2019, with 43 participants who completed the experiment. We ran the main experiment in September 2019 with 147 participants completing it. Subjects started with a tutorial and description to familiarize with the task. After the tutorial, they had the opportunity to sign up. Every participant who signed up completed one session on that same day and a second session two days later.

2.3.1.1 Payments On average participants earned ≈ $15, broken down as follows:

- $1.50 for the tutorial
- $1.00 for completing session 1
- $10.00 for coming back and completing session 2 (and thus the study)
- ∼ $3.00 in bonus payments based on choices to do additional work

We paid all subjects within three days after session 2, providing all payments at once, even if subjects dropped out early. Thus dropping out of the study did not lead to early payments. For similar reasons, subjects who completed extra work in the first session did not receive the extra bonus payment unless they also completed session 2. Instead, they received a flat bonus of $1.00 if they completed the first session, independent of the choices they made. In this way, we rule out that subjects who think that they may not come back for session 2 choose to do the extra work today in order to get extra money they would not get if they chose to work in the future and failed to come back.

2.3.2 Timeline

Table 2.1 shows the timeline of the experiment.

Table 2.1: Timeline of the first experiment

	Session 0	**Session 1**	**Session 2**	**Payment**
		(right after Session 0)	(2 days after session 1)	(within 2 days after Session 2)
Tasks	warm-up rounds of effort task	effort choices	effort task	
	consent	effort task		
	debrief survey			

2.3.2.1 The Tutorial Session 0 was the tutorial[7] which described the study and required subjects to try the experimental task to familiarize with it. As is common on MTurk, there are many Workers who sign up but don't complete the study. By only signing subjects up after the tutorial, we reduced attrition in the crucial part of our experiment substantially. The experimental task consisted of correctly counting the 1's in a matrix of

[7] This was the MTurk HIT advertised on MTurk. HIT is the acronym for Human Intelligence Task, which is one job to complete by a Worker on the MTurk platform.

1's and 0's; a task that takes between 30-60 seconds for most people. When someone gave a wrong answer, we provided a new matrix, to avoid guessing repeatedly to get the right answer.[8]

There were two sizes of matrices in the experiment, a small one (7x12 cells), and a large one (10x15 cells), and subjects completed three of each in the tutorial. Figure 2.1 shows an example of a large matrix. At the end of the HIT, Workers completed a debrief survey about clarity of our instructions, how unpleasant they found both types of matrices, and how much time the spent working and how much they earned on MTurk per week. We also elicited a survey measure of patience, used by Falk et al. (2018)[9]. Every Worker who completed the HIT, received $1.50 even if they didn't sign up for the study. Those who signed up could start the experiment right away.

0	1	1	1	1	0	1	0	0	1	1	0	0	1	0
1	0	1	0	0	0	1	1	1	0	0	0	1	1	1
1	0	1	1	1	1	0	0	1	1	0	1	1	0	1
0	0	0	0	1	0	1	1	1	1	1	0	1	0	0
0	0	1	0	0	0	0	0	1	1	0	1	0	0	1
1	0	1	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	0	0	1	0
1	1	0	1	0	0	1	0	0	0	1	1	1	0	1
0	1	0	0	0	0	1	0	1	0	0	1	1	1	0
0	0	0	0	0	1	0	0	0	1	0	0	1	0	0

Figure 2.1: Example of a large matrix

2.3.2.2 The Main Sessions

Those Workers who decided to join the study received detailed instructions, followed by a comprehension check.[10] Participants could only move on after giving correct answers. [11] Conditional on this, they started the actual study. In both sessions 1 and 2, participants had to count 25 matrices. The choices they made determined how many of these 25 matrices were small matrices (7 by 12) and how many of them were large ones (10 by 15). They made their choices using price lists. Figure 2.2 shows an example of a price list where subjects choose between 15 large matrices in two days with 90% probability for a $0.60 bonus or 22 large matrices in two days for sure for a bonus ranging from $0.00 to $3.40, increasing in $0.20 increments.

[8]Additionally to this, we allowed subjects to only get a certain number of matrices wrong, to avoid repeatedly entering the same number until a matrix pops up that has this number of 1's. We observed one subject in our pilot who we think followed this strategy and we wanted to avoid it.

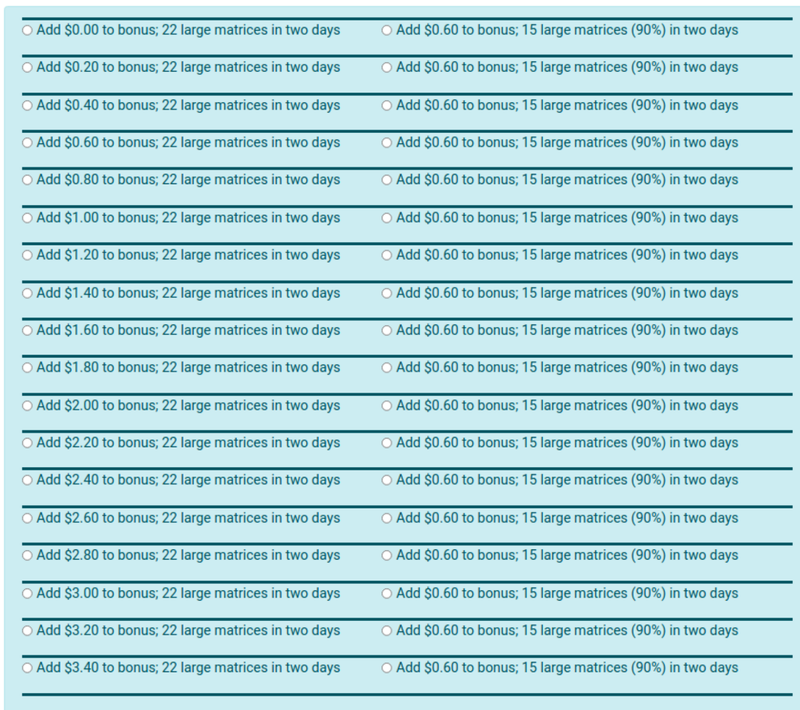
[9]The question is "How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? 0 means not willing at all, 10 means very willing"

[10]See the instructions participants saw in the Appendix.

[11]We highlighted the wrong answers, so all subjects could get them right with enough tries.

10.14754/CEU.2022.08

Participants faced the same 15 price lists in a random order (randomized for every participant) and they knew that we'd pick one choice from one of the price lists at random. If the implemented choice involved uncertainty (e.g., "20 large matrices in Session 2 with 90% probability"), then we resolved the uncertainty on the day the work was potentially due, right before subjects had to do the work. Participants then completed the work for session 1, after which we gave them the link to Session 2 that opened in two days. We also sent a reminder email on the day of session 2 for them to come back. In session 1, we additionally asked participants after they made their choices how they went about making these choices.



Add $0.00 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $0.20 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $0.40 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $0.60 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $0.80 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $1.00 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $1.20 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $1.40 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $1.60 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $1.80 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $2.00 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $2.20 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $2.40 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $2.60 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $2.80 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $3.00 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $3.20 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days
Add $3.40 to bonus; 22 large matrices in two days	Add $0.60 to bonus; 15 large matrices (90%) in two days

Figure 2.2: Example of a price list

2.3.2.3 Choosing large matrices rather than extra matrices We decided to let participants choose the number of hard tasks, rather than the number of extra tasks to avoid making them choose primarily based on the extra time taken and get them to think more about the difference in unpleasantness. The primary reason for doing this was that our subjects spent a lot of time working on MTurk (roughly 20 hours per week), and may have developed heuristics based on the time it takes to do a task. Even more problematic is if MTurkers consider our task to be roughly as tedious as other tasks on MTurk, in which case they would primarily decide based on whether we pay more or less per hour. In that case, a choice to work less today might be driven not by present bias but by the fact that that the hourly wage offered for today is lower than the hourly wage offered in the future, given that the worker may already have decided to work several hours each of those days. In order to make such thinking less likely, we decided to let subjects choose the number of large tasks, rather than extra tasks.

2.3.3 Implementation of risk as an excuse

We test for excuse-driven present bias in two batches of four choices.[12] In each batch, we use the switching point in a price list as the indifference point, which gives us 4 inferred indifference points corresponding to X, X', Y, and Y' as described in section @ref(edpb). Our pre-registered alternative hypothesis was that $\Delta\Delta = (Y' - Y) - (X' - X) > 0$. For example, the $\Delta\Delta$ we get in batch 2, with the excuse option being 15 large matrices in the future with $p = 0.9$, is the following:

$$\Delta\Delta = 0.9d(15) \cdot (\beta - \beta_E) \tag{2.2}$$

(see the exact choices in the Appendix).

Based on Exley (2016), our hypothesis was that in the excuse condition, people's willingness to work in the present decreases more than the willingness to work in the future, leading to $\beta_E < \beta$, and an observed positive $\Delta\Delta$ on average.

2.4 First experiment results

We asked participants to rate the large and small tasks on a 10-point scale, comparing them to other tasks on MTurk. Participants on average rated the large matrix as significantly less pleasant than the small matrix (3.43 vs 4.85, 5.0 = equally pleasant to other MTurk tasks), although some participants stated clearly that they didn't mind in their description of how they chose.[13] Out of the 154 participants who completed Session 1, 147 also completed session 2 and thus the experiment. Therefore, between the two sessions, attrition was only 4.5%. For the analysis of choices, we use the sample of those who completed all aspects of the experiment, to exclude those who, at the point of making choices, might have already known they would not finish it all. Only 1.7% of the choices had multiple switching points. We excluded these choices and the other choices in that batch from the analysis, as we could not infer an indifference point from them.[14] We present the results for our main hypothesis in this section. First, we show evidence for excuse-driven present bias in batches 1 and 2. Then, as a robustness check, we show that the effect we find cannot be driven by concave utility over money.

[12] 8 of the price lists correspond to the 2 batches, 4 choices correspond to a third batch – in which we later realized that we made a mistake in the options, so identification is not possible in that batch –, and the remaining 3 price lists provide extra data for identifying β and direct choices between no-excuse and excuse alternatives.

[13] We asked this question at the end of Session 0, in a debrief survey. See other debrief survey statistics in the Appendix.

[14] As our outcome of interest, $\Delta\Delta$, is a difference in the differences of such indifference points, we had to exclude all four choices in a batch if there was at least one choice with multiple switching points.

10.14754/CEU.2022.08

2.4.1 Main results

Figure 2.3 shows the distribution of the excuse-driven increase in the willingness to work in two days over willingness to work today ($\Delta\Delta$). Our main pre-registered test is a two-sided t-test on $\Delta\Delta$ for batches 1 and 2 jointly (although we report the individual t-tests too).[15] Specifically, we run the following regression, separately for each batch b, and then for batches 1 and 2 combined:

$$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b} \tag{2.3}$$

Figure 2.3: Excuse-driven difference in WTW in two days over WTW today

Table 2.2 shows the results for this simple regression. Columns 1 and 2 correspond to the batches separately, while in column 3 the two batches are pooled together. In column 3, standard errors are clustered at the participant level, since we have 2 observations per individual, one for each batch. The results show that there is an \$0.11 increase in the difference between willingness to work in two days vs today in the excuse condition. This is roughly 0.2 standard deviations of the difference between willingness to work in two days vs today. As described in Section 2.2, state-independent risk preferences cannot explain the result, nor are we aware of any existing theory about risk and time interactions that would explain it. One possible caveat could be concave utility over money. We rule out this explanation in the next section.

2.4.2 Ruling out concavity of money

Concave utility over money might also lead to $\Delta\Delta > 0$ depending on the choices offered.[16] Suppose that for a participant, $Y' = 9$, $Y = 6$, $X' = 2$, and $X = 0$. Then, $Y' - Y = 3$ and

[15] Our hypothesis suggests a one-sided t-test, however we decided against pre-registering it as such since one-sided tests tend to be frowned upon. Alternatively our test can be interpreted as a one-sided t-test with 2.5% significance level.

[16] Over these small stakes, concave utility over money might be the result of loss aversion, or various types of framing.

Table 2.2: Results: t-tests

	$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b}$		
	Batch 1	Batch 2	Batches 1&2
	(1)	(2)	(3)
Constant	0.113*	0.106**	0.110**
	(0.062)	(0.053)	(0.043)
Observations	138	141	279

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses,
Batches 1&2: clustered at the individual level

$X' - X = 2$, which would lead to $\Delta\Delta = 1 > 0$. However, with concave utility over money the increase from \$0 to \$2 and from \$6 to \$9 could be the same. To rule this out, we chose the payments such that for most participants we expected $X \geq Y$. In this way, concavity of money would if anything work *against* us finding an effect, pushing $\Delta\Delta$ down. Table 2.3 shows the results for a restricted sample of those for whom $X \geq Y$, so for whom the results cannot be explained by concave utility over money. The results for these participants are even stronger; for them, the excuse-driven increase in the willingness to work in two days over work today is \$0.19.[17]

Table 2.3: T-test equivalents, ruling out concavity of money

	$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b}$	
	Batches 1&2, all	Batches 1&2, $X \geq Y$
	(1)	(2)
Constant	0.110**	0.190***
	(0.043)	(0.049)
Observations	279	214

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses,
clustered at the individual level

Although we are not aware of any such theory, some form of risk and time interactions with state-dependence (other than excuse-making) might explain our results. To overcome these issues, we planned a follow-up study in a laboratory with students. We turn to the description of this lab experiment in the next section.

[17] We have no intuition for why the effect is stronger on this subset.

2.5 Second experiment

We decided to repeat our experiment with students in a lab, due to several drawbacks of MTurk. To measure time preferences, we need a task or a consumption good that participants cannot substitute easily with other tasks they perform regularly. The MTurk Workers in our sample report working on average 20 hours weekly on MTurk, on tasks that may be close substitutes to our tasks. If our tasks are as unpleasant as other MTurk tasks, MTurkers may choose based on which tasks pay more, including whether we pay more per task in session 1 or session 2. Our participants find the small matrix similarly pleasant to other MTurk tasks (4.85 on a 0-10 scale, 5 = equally pleasant), while they considered the large matrix to be less pleasant (3.43), which suggests that we partly solved this issue with the two sizes of matrices. However the difference in pleasantness is not huge, and as we mentioned earlier, many subjects reported not caring differently about the small or large matrices. We supposed that students in the lab were less likely to substitute our tasks for equally tedious and paid tasks, which makes them a better subject pool for eliciting time preferences.

2.5.1 A chance of no extra tasks as an excuse

Since we assumed that students do not substitute our tasks with other paid tasks outside of our experiment, we decided to modify the risk excuse slightly. In each session, our subjects had to complete 10 required tasks, and the options they could choose from involved having to do *extra* tasks for bonus payments. Therefore the no-excuse alternative meant having to do extra future tasks on top of the 10 required ones for sure, while the excuse alternative introduced a 10% chance of not having to do the extra tasks in the future (see all choices in the Appendix). The identification strategy presented in Section 2.2 applies here, as well.

2.5.2 A different type of task as an excuse

We additionally wanted to strengthen our findings with other potential excuses, rather than only for risk. Our idea was that we introduce two different types of tasks that are commonly used as effort tasks in the experimental literature – the matrix counting task, and a blurry Greek letter transcription task (see an example in Figure 2.4).[18] We then use a similar difference-in-differences strategy as with risk to identify excuse-making, using four choices. In the no-excuse condition people make choices about a *baseline* task, say, the matrix counting task. In the excuse condition, they have the option to choose another type of task – the blurry Greek letters task – for future work.

No-excuse condition:

$$\begin{aligned} x_1 \text{ matrices in the future} + \$X &\sim x_2 \text{ matrices in the future} \\ x_1 \text{ matrices today} + \$X' &\sim x_2 \text{ matrices in the future} \end{aligned}$$

Excuse condition:

$$\begin{aligned} x_1 \text{ matrices in the future} + \$Y &\sim x_2 \text{ Greek tasks in the future} \\ x_1 \text{ matrices today} + \$Y' &\sim x_2 \text{ Greek tasks in the future} \end{aligned}$$

[18] This task was used e.g. in Augenblick and Rabin (2019).

10.14754/CEU.2022.08

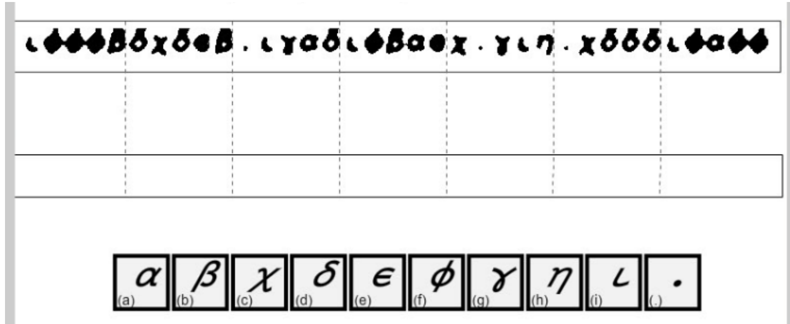


Figure 2.4: Example of a Greek transcription task

The excuse-driven change in the willingness to do x_1 matrices in the future over today is then, analogously to our identification with risk in Section 2.2:

$$\Delta\Delta = \Delta_T - \Delta_C = d_G(x_2) \cdot (\beta - \beta_E) \tag{2.4}$$

where the disutility of doing x Greek tasks is denoted with $d_G(x)$. The idea is that if people want to choose to work in the future, they can rationalize their choice by saying that they chose that option because it offered a different task. However, our design, by using the asymmetry in the choices, allows to distinguish just preferring the other type of task from indeed using the other task as an excuse to choose to work in the future.[19]

2.5.3 Technical details

We pre-registered the second experiment the AEA RCT Registry as well (Drucker and Kaufmann, 2019). The experiment ran between May and June 2020 with student subjects of the lab of the Luxembourg Institute of Socio-Economic Research (LISER-LAB). Due to the restrictions because of the COVID-19 pandemic, the experiment was conducted entirely online. It ran over four weeks with two sessions per week, on Mondays and Thursdays. We moved from two days between sessions to three days to allow for a lower estimated β, as Augenblick (2018) finds a few hours' β of 0.94, a daily β of 0.91 and a weekly one of 0.87. We chose to run the experiment for four weeks to gain more data points, and to let participants become familiar with the setting. Participants received a €15 completion bonus each week if completing both sessions that week, but we did not require them to participate in all weeks; a participant could complete one week then skip the next one and come back the third week, for example. There was an extra €15 bonus for completing all four weeks. The total payment if a student participated in all sessions throughout four weeks was around €95. Four waves of students started the experiment, each wave one week after the previous one. In total, 75 students started the study and 47 stayed until the end.

Similarly to the first experiment, participants made all decisions using price lists. Four price lists added up to a batch to identify $\Delta\Delta$ – the extent of excuse-making in present

[19] Hiding the choice of future tasks behind the choice of a different type of task is close to the implicit preferences framework presented in Cunningham and de Quidt (2022).

bias. In each week we asked four risk batches and four matrix-versus-Greek batches. All four questions of a risk batch involved the same task – Greek or matrix – but the type of task was randomized within participant across batches. In the matrix-versus-Greek batches one task was the baseline and the other one the excuse task, also randomized across batches. Each Thursday, participants made work decisions for two weeks ahead, to elicit long-term preferences, and each Monday, they could revise their decisions for that week, to elicit short-term preferences. This allowed us to differentiate between present bias and exponential time discounting (as in Augenblick et al., 2015). See Table 2.4 for an illustration of the timeline of the experiment.

Table 2.4: Timeline of the second experiment

	Week 1	**Week 2**	**Week 3**	**Week 4**	**Payment**
Session 1	Instructions,				
(Monday)	consent				
	effort choices	effort choices	effort choices	effort choices	
	for Week 1	for Week 2	for Week 3	for Week 4	
	effort task	effort task	effort task	effort task	
Session 2	effort choices	effort choices			
(Thursday)	for Week 3	for Week 4			
	effort task	effort task	effort task	effort task	

2.6 Second experiment results

In the second experiment we used the same simple regression as in the first one:

$$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b} \tag{2.5}$$

We report the results jointly for all batches within an excuse type. Table 2.5 shows the four main estimated $\Delta\Delta$-s. The first column shows the estimated $\Delta\Delta$ in the short-term work decisions made for that week when the excuse was risk. The second column shows the long-term $\Delta\Delta$ in the work decisions for two weeks ahead. The third and fourth columns report the estimated $\Delta\Delta$-s in the decisions where the excuse was a different type of task (matrix-versus-Greek). The third column is the short-term $\Delta\Delta$ and the fourth the long-term $\Delta\Delta$ of this excuse type. We expected the first and third $\Delta\Delta$-s to be significant and positive, and the second and fourth to be zero or smaller positive. Neither of the $\Delta\Delta$-s turned out to be significant, and the signs are not always as expected, either.

In the first session, we asked the participants to fill in a survey about some aspects of the study. Two of the survey questions were to rate on a 0 to 10 scale how pleasant they find each type of task. From the two ratings we can see which task they prefer. Since the measured $\Delta\Delta$ depends on the disutility of the task, it might be possible that there is an excuse-driven present bias for the disliked task but not for the preferred task. In the matrix-vs-Greek decisions, it is possible that the disliked task is not a good excuse

Table 2.5: Main results

	Dependent variable:			
	ΔΔ			
	Risk short-term	Risk long-term	M-v-G short-term	M-v-G long-term
	(1)	(2)	(3)	(4)
Constant	0.060	-0.086	-0.032	0.105
	(0.066)	(0.076)	(0.068)	(0.087)
Observations	411	293	259	166

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the individual level.
Clustered standard errors in parentheses.

to postpone the preferred task, but the excuse could work the other way around. Table 2.6 looks at whether we get different results in the short-term risk decisions by whether the task is the preferred or the disliked one, and in the matrix-versus-Greek decisions by whether the baseline task is preferred or disliked. None of the ΔΔ-s are significant with this distinction, either.

Based on these results we think that the different type of task does not work as an excuse. Although, a large majority, 63 percent of the participants preferred the Greek task over the matrix task, and only 17 percent preferred the matrix task, while 20 percent were indifferent. It is possible that with much lower differences in unpleasantness between the two tasks, the excuse of another task could have worked better. In any case, because of the promising MTurk results, we still think that risk could work as an excuse.

Table 2.6: Short-term ΔΔ by preferred/disliked task

	Dependent variable:			
	ΔΔ			
	Risk preferred	Risk disliked	M-v-G preferred	M-v-G disliked
	(1)	(2)	(3)	(4)
Constant	0.041	0.069	0.088	-0.159
	(0.086)	(0.087)	(0.080)	(0.103)
Observations	164	169	96	111

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the individual level.
Clustered standard errors in parentheses.

Although students are not doing these types of tasks often, it is still possible that their choices do not only depend on the disutility of our tasks and their time preferences, but they also substitute our tasks with other tasks unrelated to the experiment. To test whether they make different decisions in questions similar to the MTurk study, we recruited 30 new participants at the end of the second experiment. This follow-up was not pre-registered. In the follow-up we reintroduced the choices over hard vs easy tasks instead of extra tasks. We also made the choices more simple,[20] to minimize the possibility that participants decide randomly. These participants made four batches of risk choices, only with matrix tasks (see the exact choices in the Appendix). Table 2.7 shows the $\Delta\Delta$ for these simpler, intensive margin choices. There is a large positive $\Delta\Delta$ for these choices that is significant at the 10 percent level:[21] The excuse-driven increase in the willingness to work in three days over work today is €0.24.

Table 2.7: Follow-up risk choices

	Dependent variable:
	$\Delta\Delta$ Risk short-term
Constant	0.242* (0.146)
Observations	113
Note:	*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses.

From these results we suspect that different mechanisms work when choosing to do extra work or hard work. In some of the batches participants had to choose between the same amount of work today or in three days, so from these batches we have a reduced-form measure for present bias. From the choices between the same amount of work for sure in three days and with 90% probability in three days we can estimate a reduced-form measure of the value of the 10% drop in probability. Table 2.8 shows these values in the extra-work and the hard-work choices, only at the risk batches. Although we only had few such choices, it is clear that participants exhibit some present bias in both decisions: they ask for 72 cents more for doing the same amount of extra tasks today instead of in three days, and 41 cents more for changing the same amount of tasks from easy to hard today instead of in three days. Unfortunately, we did not ask these questions two weeks into the future, so we do not have such reduced-form measures for δ only. In the third and fourth columns we can see the value of the 10% drop in the probability of future tasks. Participants clearly value the drop – therefore the excuse – positively: they ask for 48 cents more for extra future tasks for sure instead of extra future tasks with 90% chance, and 47 cents more for hard instead of easy tasks for sure instead of the same option with 90% probability.

Though there is present bias and a positive valuation of the excuse option in both types of decisions, Table 2.9 suggests that the estimated excuse-driven present bias in

[20] E.g., 10 tasks today vs. 10 tasks in 3 days, instead of 12 tasks today vs. 8 tasks in 3 days.

[21] Since we have only 30 participants, the standard errors are not clustered.

Table 2.8: Reduced-form direction of $\beta\delta$ and value of the 10% drop

	Dependent variable:			
	$(1-\beta\delta)d(x)$		$(1-0.9)\beta d(x)$	
	Extensive	Intensive	Extensive	Intensive
	(1)	(2)	(3)	(4)
Constant	0.717***	0.407***	0.476***	0.469***
	(0.181)	(0.120)	(0.124)	(0.150)
Observations	41	57	42	58
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Standard errors in parentheses.			

these choices might be very different. In this experiment we do not have enough power to compare the two types of choices, so we plan to run a third experiment to test whether there is indeed an excuse-driven present bias in the hard-work choices but not in the extra-work choices. If the hard-work results replicate but we find no effect for the extra-work choices, that would suggest that there are two different underlying mechanisms: In the extra-work choices, more work crowds out other tasks with similar disutility but lower pay, so the choices rather hinge on hourly wage instead of disutility of the task. In the hard-work choices, on the other hand, harder tasks replace easier tasks with lower pay, so the choices hinge on disutility of the task, hence on time preferences.

Table 2.9: $\Delta\Delta$

	Dependent variable:	
	$\Delta\Delta$	
	Risk extensive	Risk intensive
	(1)	(2)
Constant	-0.000	0.439*
	(0.127)	(0.220)
Observations	41	57
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses.	

The other issue with our second experiment results is power. In the end, we only had 75 participants and 5-6 data points per participant on average from the short-term risk choices, even less from the long-term ones. With this number of observations we had 68 percent power to detect a $\Delta\Delta$ of 11 cents. To have 80 percent power, with 5 data points per person we would have needed at least around 100 people, or the same amount of people but 10 data points per person. If we look at the short-term risk results separately by task (Table 2.10), we get a $\Delta\Delta$ of similar magnitude for the matrix task to that in the MTurk experiment, but a close to zero $\Delta\Delta$ for the Greek task, both insignificant. Therefore, we plan the final experiment to last only one week, to reduce attrition, and to have enough observations for 80 percent power to detect a $\Delta\Delta$ of similar magnitude. To be able to separate short-term excuse-driven behavior from long-term, we are going to have three sessions: a session 0 to make effort choices for Sessions 1 and 2, and 2 work sessions. In Session 1, participants will be able to revise their effort choices. We aim to test both hard-work and extra-work choices, to look at whether these choices indeed depend on different factors.

Table 2.10: Short-term $\Delta\Delta$ by type of task

	Dependent variable:	
	$\Delta\Delta$	
	Risk matrix	Risk Greek
	(1)	(2)
Constant	0.105	0.012
	(0.079)	(0.097)
Observations	212	199
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard errors clustered at the individual level.
		Clustered standard errors in parentheses.

2.7 Discussion

We tested whether people behave in a more present-biased way if they have excuses to do so. Our two experiments yielded mixed results. Still, the MTurk and the LISER-LAB results on risk suggest that a chance of no work/ easier work in the future makes people asymmetrically make different intertemporal work decisions than when all work is certain. We call this phenomenon excuse-driven present bias, but other mechanisms can work in the same direction, too. There is evidence that risk and time preferences are intertwined. E.g., people might exhibit present biased behavior because they think that future consumption is uncertain (Chakraborty et al., 2020), or they might trade off the probability and timing of consumption (Baucells and Heukamp, 2012). Interrelated time and risk preferences would only explain our results if they were context-dependent: we identify excuse-driven present bias from a change in the valuation of an option depending on the alternative. We are unaware of such a theory that would predict participants treating risky future disutility differently by the alternative options.

It is possible that our participants have reference-dependent preferences (Kahneman and Tversky, 1979; Kőszegi and Rabin, 2006) for the tasks, and they always set the fixed option in the price list as their reference point. In this case, if the excuse option is worse in itself than the no-excuse option, reference dependence also predicts a positive $\Delta\Delta$. However, if we set the excuse option to be better than the no-excuse option for most subjects, reference dependence works against us, predicting a negative $\Delta\Delta$. We made sure in our batches that the excuse option is at least as good as the no-excuse option, so if we find a positive $\Delta\Delta$, that is *despite* potentially reference-dependent preferences (see a formal derivation in Appendix Section B.3.4). In fact, the signs of the coefficients in Table 2.6 in the matrix vs. Greek choices can indicate reference-dependent preferences. It is also possible that the positive $\Delta\Delta$ we measure is a consequence of a framing effect (Tversky and Kahneman, 1981) or attention (Simon, 1971). Still, reacting differently to a particular framing of a choice or differential attention to details of a choice depending on the alternative is really similar to a mechanism that we call excuse-making.

The two excuses we test are different in a sense that one allows the person to hide her present bias behind another dimension of the choice: choosing a different task, while the risk excuse rather rationalizes the choice of future work by allowing avoiding it altogether. We think excuses can work both ways. People might use risk unconsciously for postponing work in the hope of not having to do it in the end. A different type of task is an explanation that they might give to others when asked why they chose to postpone work.

At the end of our second experiment, we asked participants to give their opinion on some topics related to procrastination and excuse-making. They had to indicate on a 0 to 10 scale how much they agree with different statements. The results are shown in Table 2.11. The first two questions aimed to elicit whether students think they are prone to some typical procrastination behaviors: ending up working out too little or always postponing household chores. On average, our students neither agree nor disagree with the corresponding statements. They are also indifferent in the question of whether they make excuses when postponing work to do. However, they rather agree with the statement that people in general make excuses for postponing work, and with people postponing work in the hope of not having to do it in the end. Based on these answers and the results of both experiments on the risk excuse, we hope to be able to corroborate in a third, final experiment that there can be excuse-making in present bias.

Table 2.11: Final survey results

Statistic	N	Mean	St. Dev.	Min	Max
I usually postpone household chores.	74	5.027	2.998	0	10
I usually work out too little.	74	5.392	3.355	0	10
People usually make excuses for postponing work they have to do.	67	7.463	1.995	1	10
I usually make excuses for postponing work I have to do.	67	4.881	2.826	0	10
People postpone their work in the hope of not having to do it later.	67	6.149	2.624	1	10

10.14754/CEU.2022.08

3 Chapter 3: Compensatory Advantage and Inequality in Educational Aspirations

3.1 Introduction

There is high correlation between the education levels of parents and their children. In 2012, 52% of 25-32 year-olds had the same education level as their parents on average in the OECD countries (OECD, 2015). Parental education influences children's education outcomes through several channels. First, parental background affects cognitive and non-cognitive skills of children already in early ages. Parental investment in early childhood is crucial (see a review in Heckman and Mosso, 2014). Falk et al. (2019) find that children from high socioeconomic background are more intelligent, more altruistic, less risk-seeking, and more patient already at ages 7-9. The main channels through which parental social status affects children's IQ and preferences in early childhood are time spent with the children and parenting style. Whether these differences remain or diminish depends largely on the education system. School systems which track students into academic or vocational tracks later can increase intergenerational educational mobility by decreasing inequality in the school performance of students (see e.g., Pekkarinen, 2018; Schütz et al., 2008). Besides entering school with different skills, children from low and high socioeconomic status (SES) also differ in their resources to cope with hardships during their educational career. A strand of literature in sociology explores a specific channel of low intergenerational educational mobility: the compensatory advantage of high socioeconomic status. After a negative school event – like failing a subject or a bad school choice –, high-SES students have much more resources to compensate for this negative shock, and they recover from the shock faster than low-SES students. This compensatory advantage reinforces the initial differences in socioeconomic status (Bernardi, 2014; Bernardi and Cebolla-Boado, 2013; Bernardi and Grätz, 2015).

I am addressing the compensatory advantage channel, by looking at how the educational aspirations of children from different socioeconomic backgrounds change after having to repeat a grade at the end of their primary school career. As more and more difficult subjects come in higher grades of primary school, students who had difficulties earlier might have an even harder time in these higher grades. Grade retention is originally there to give a second chance to students who failed one or more subjects for the first time. Therefore, it could be a way to decrease the inequality in the performance of students, and help worse students achieve higher education levels than if they were promoted to the next grade without the necessary qualifications. On the other hand, grade retention is often associated with a stigma, which makes it more difficult for the retained student to catch up. The existing evidence on the effects of grade retention in later years of primary school is inconclusive. Jacob and Lefgren (2004) and Jacob and Lefgren (2009) use a regression discontinuity design and a standardized test-based promotion system in the US to look at short-run and long-run effects of grade retention. They find that retention does not have a consistent effect in the short run, and retaining students in the 6^{th} grade of primary school does not have long-run effects – on high school completion–, either. However, retention in 8^{th} grade increases the probability of dropping out of high school. Gary-Bobo et al. (2016) develop a multi-stage human capital accumulation model to look at the effects of grade

10.14754/CEU.2022.08

retention in grades 6 to 8 (first three grades of junior high school in France) on 9^{th} (final) grade outcomes. They find a small positive average treatment effect on the treated on test scores, and the effect is higher for lower performer students. However, grade retention decreases the probability of entering 9^{th} grade for all students.

Using rich administrative and survey data from Hungary, I look at grade retention in the 7^{th} grade of primary school, and estimate the differences in educational aspirations in 8^{th} grade between retained and non-retained children conditional on their socioeconomic status. I also look at how the changes in aspirations translate to the type of secondary school track they are in in 10^{th} grade. Though it is important to know how grade retention affects track choice and educational attainment, we know little about the mechanism through which retaining a student leads to these outcomes. Are these children simply not able to catch up, which leads to lower educational outcomes? Or they also lose their confidence and set lower goals which is why their educational outcomes will be worse than those of non-repeaters? To answer these questions, we have to look at how children adjust their aspirations after retention. To my best knowledge, existing papers only study the effect of early grade retention on aspirations. Hughes et al. (2013) find that parental expectations about their child's highest education level decrease after retention in the 1^{st} grade of primary school. They find that decreasing parental expectations play a role in the negative effect of retention on 3^{rd} grade performance of children. Cham et al. (2015) look at the effect of grade retention in the primary grades on students' own expectations in 9^{th} grade about finishing high school. They use propensity score weighting to equate the distribution of pre-treatment characteristics of retained and non-retained students. They do not find any effect of grade retention on motivation to finish secondary school in 9^{th} grade. They find, however, that retained students in 9^{th} grade value a high school diploma more, and feel that their teachers and peers are more likely to expect them to graduate.

Aspirations act as reference points that induce motivation through loss aversion (Heath et al., 1999; Page et al., 2007). These reference points are highly influenced by one's social environment and so, through the individual actions motivated by them, may reinforce economic inequalities (Genicot and Ray, 2017, 2020). If aspirations are so important in what people achieve, it is worth looking at how they are shaped by negative events during one's educational career. My main outcome of interest is the educational aspirations at 8^{th} grade. I proxy socioeconomic status with parental education. In my main specifications, I define low SES as both parents having lower qualification than a high school diploma, and high SES as at least one parent having a high school diploma. Educational hardships might cause a decline in low-SES students' dream education level, while high-SES students might stay focused on the education level they wanted initially to achieve. I find evidence for heterogeneities in the aspirations and the paths of students after retention.

Low-SES repeaters are 15.6 percentage points more likely to aspire for at most a secondary vocational certificate in 8^{th} grade than non-repeaters, while the share of high-SES wanting to achieve this level does not differ by retained status, controlling for 6^{th} grade aspirations and other characteristics. While already in 6^{th} grade 24.8 percent of low-SES children want to achieve at most a secondary vocational certificate compared to 3.8 percent of high-SES, the SES gap is even larger for retained students. I also look at the aspiration for tertiary education. Here, low-SES repeaters do not change their (already quite low) tertiary aspirations, but for high-SES repeaters, the difference is -7.9 percentage points. Therefore, the SES-gap in tertiary aspirations is actually lower for retained than non-retained students. When looking at the effect of retention on the aspired years of education by pre-retention aspirations, I find that the compensatory advantage of high SES is highest at high levels of initial aspirations: while low-SES children decrease their aspirations significantly, high-SES with initially high aspirations stay on the same track.

In Hungary, a student can be retained if she fails at least one subject in school or if she was missing from school for a significant amount of time and could not pass a grading exam. When controlling for characteristics that might have led to the student failing a subject or missing from school, high-SES children are still less likely to get retained than low-SES children. 6^{th} grade midterm mathematics performance in school seems to be one factor that, if the student performs poorly, affects students' 7^{th} grade retention similarly, regardless of socioeconomic status. When splitting the sample by 6^{th} year midterm mathematics grades, the compensatory advantage disappears for the lowest-performing students but it is there for higher performing ones, where retention is much rarer. Using a smaller, administrative database and controlling for factors that are good proxies for missing from school because of health issues – visits to the general practitioner and days spent in a hospital in 7^{th} grade –, the results stay similar, though they become much weaker because of the low sample size. Finally, I look at the differences by socioeconomic status within these groups in transitioning to a secondary school track that gives access to tertiary education. The patterns here are similar to the aspirations: low-performing students are 11 percentage points less likely to get into this track if retained, but there is no compensatory advantage of high SES among these students. There is, on the other hand, a compensatory advantage for the higher performer students: low-SES repeaters are 6-24 percentage points less likely to attend this institution, while high-SES repeaters are not less likely to attend than non-repeaters. When controlling for health characteristics in the administrative database, on the other hand, all coefficients become smaller and insignificant, and the compensatory advantage at each mathematics performance level disappears. We have to treat these results with caution, though, because the numbers of repeaters in these samples are very small.

3.2 Data

I use two datasets for this paper. For most of the descriptive statistics and the baseline regressions I use the National Assessment of Basic Competencies (National ABC) database that contains rich, administrative and survey, educational data of all Hungarian students in the 6^{th}, 8^{th}, and 10^{th} grades, between 2008 and 2017. For learning more about the reasons behind grade retention I use an administrative dataset that covers half of the Hungarian population of ages 5-74 between 2003 and 2017. This dataset contains demographic and labor market data, along with healthcare-related variables, and can also be linked to the National ABC dataset.

3.2.1 National Assessment of Basic Competencies data

The first data source I use is the National ABC database that covers the period between 2008 and 2017. The National ABC is a standardized mathematics and reading comprehension test that all Hungarian students in 6^{th}, 8^{th}, and 10^{th} grade of public education take. The test has been conducted yearly since the 2005/2006 schoolyear. Compulsory education in Hungary for students in the sample started at age 6 or 7 (depending on the birth month of the student), so the tests are taken by 12/13, 14/15 and 16/17 year-olds. All students have to write the test, except those with autism and intellectual disabilities.The test is centralized, administered by the Education Authority, and aims at measuring the problem solving skills of students in mathematics and reading comprehension, rather than knowledge of school material. Students take the test in their own school, at the end of May. From 2008, students are identified by a unique identifier, the so-called OM code,[22] so 6^{th}, 8^{th}, and 10^{th} grade data of the same students can be linked. The National ABC database contains the standardized test scores and various background characteristics of students and their families from a background questionnaire.

3.2.1.1 Background questionnaire The tests are complemented by three background questionnaires on student, institution, and, if a school has multiple branches, branch level. Students complete the student questionnaire on paper at home, and the headmaster of the institution, and, in case of multiple branches, the head teacher of the branch completes the institution and the branch level questionnaires online. Completing either questionnaire is voluntary. The student questionnaire contains 47 questions, and students take it home and fill it in with the help of their parents. I only use variables from the student questionnaire, which contains variables regarding the student's academic progress (last year GPA, last midterm GPA and grades from main subjects, number of years in kindergarten, grade retention in different phases of the educational career, educational aspirations, how much they like specific subjects, extracurricular activities), family background (status on regular child protection allowance, subsidized meals, family members living with the student, parents' age, education level, and labor market status), household characteristics (size of household, age composition, number of rooms, books, bathrooms, computers, internet access, etc.), family activities, and the student's perception of the wealth of the family compared to neighbors.

3.2.2 Administrative database

The second database I use is an administrative dataset (Admin3) that contains rich data from half of the Hungarian population of ages 5-74 between 2003 and 2017 (see a detailed description in Sebők, 2019). The main file of the Admin3 contains demographic data, labor market status, income, job characteristics, social transfers, and education status of each person in the dataset, on a monthly level. This main dataset is then linked by an individual ID to administrative health care data provided by the National Health Insurance Fund (*Nemzeti Egészségbiztosítási Alapkezelő* - NEAK), and administrative and survey-based educational data from the National ABC database.

3.2.2.1 Health care data For better predicting retention I use administrative health care data from the Admin3 that comes from the National Health Insurance Fund. This

[22]OM is the abbreviation for the Ministry of Education *(Oktatási Minisztérium)*

dataset contains monthly data of each insured person about the number of visits to the general practitioner, costs of outpatient care covered by social security, costs paid by the insured on medication, cost of purchased medication covered by social security, costs of inpatient care covered by NEAK, and whether the person is eligible for prescription exemption.

3.2.3 Definition of variables

For the analysis I have to define the variables of interest. These are the treatment variable – repeating 7^{th} grade, educational aspirations, and socioeconomic status. For the first one, children state in the questionnaire whether they had to repeat a grade once or multiple times in different parts of their school career. I use this variable to control for whether someone repeated a grade up to the 6^{th} grade. As for the treatment variable, I simply treat students as having repeated between 6^{th} and 8^{th} grades if more than two years passed between their first 6^{th} grade test and their first 8^{th} grade test. If there are more instances of the same person in the 6^{th} grade database, I regard these instances as repetitions of 6^{th} grade. From the number of 6^{th} grade observations and the years passed between the first 6^{th} and 8^{th} grade occurrences, I can deduce if a child repeated the 7^{th} grade. Since children take the National ABC test in May, by May they already suspect if they will have to repeat the grade, so 6^{th} grade aspirations partly reflect the effect of later retention. Due to this potential effect on pre-treatment variables, I use 7^{th} grade retention as the treatment and I exclude those who were retained in 6^{th} grade. There is a practice among children who apply for 6-year academic secondary schools in 6^{th} grade but do not get admitted: they complete 7^{th} grade in their primary school, apply again, and if admitted, repeat 7^{th} grade but in the new institution they were admitted to. I set the treatment to be 0 for these children (if they have not repeated the 6^{th} grade either), because their grade repetition is of a very different nature than what I am interested in.

In the questionnaire, children choose the education level they want to reach from a list of qualifications from primary school to a doctoral degree. I create two variables from this data. The first one is educational aspirations in three categories: at most vocational qualification, high school diploma or a post-secondary non-tertiary qualification, and college or above. For the second variable, I assign years of education to each qualification level in the questionnaire: primary school is 8 years, a vocational degree is 11 years, and so on (see the construction of the variable in more detail in the Data Appendix section C.0.4.2).

The last variable that is needed to define is socioeconomic status (SES). For the main specifications, I use a categorical SES variable. I proxy socioeconomic status with parental education, which is also asked in the survey in fine categories: from unfinished primary education to university degree. I create two categories: low-SES is a child whose parents do not have a high school diploma, while high SES means at least one parent has a high school diploma. I chose this distinction after Falk et al. (2019), however, they also use information on family income which unfortunately I do not have. Also, since there are very few repeaters among children of tertiary educated parents, drawing the line at a lower parental education level gives enough repeaters in both SES categories. In a robustness check I use a three-category SES variable, where high SES is further decomposed into two categories: medium SES here means at least one parent with at least a high school diploma, and high SES means at least one parent with a tertiary degree. In another robustness check, I use a continuous variable, where I apply the same rule to convert the categories to years of education as for the educational aspirations. For the continuous SES variable I take the parent with the highest education level.

3.2.3.1 Grade retention rules in Hungary Grade retention in Hungary is regulated by the 2011 Public Education Law.[23] In the first grade of primary school, parents can request grade retention for their child even if the child fulfilled all educational requirements. The school principal has to approve this request. From the second grade and above, grade retention has two sources: first, if the child receives an insufficient grade in at least one subject, she can take a grading exam in those subjects at the end of summer. If she fails, she has to repeat the entire grade. Teachers have a say in the decision about whether to fail someone in a subject or make her do some extra coursework in order to pass the subject without having to take the grading exam. The second source is absenteeism. If the child missed at least 250, she has to take equivalency tests in the subjects in which she cannot receive a final grade at the end of the school year. She has to pass all tests to be able to proceed to the next grade. If the child has at least 250 unjustified absences, she has to repeat the entire grade, without the possibility to take the equivalency test. Retention is also automatic if the child receives an insufficient grade from at least 3 subjects. In practice, grade retention in Hungary is quite rare. Figure 3.1 shows the percentage of retained students in the upper grades of primary school by year. Usually 2 to 3 percent of students are retained in these grades, and retention in the final grade of primary school is even rarer: less than 1 percent in most years. Grade retention in the lower grades of primary school is not very common either, except in the first grade, where around 3.3-4.5 percent of children were retained in my sample.

There is another reason why someone can be retained: if the family spends one or more years abroad, the child has to prove that she learned everything her peers learned during the schoolyear by passing a grading exam. Often the skills and knowledge the child learns abroad are very different from what the Hungarian school system requires and the child cannot pass the grading exam, leading to having to repeat a grade (see the experiences of returning children in Árendás et al., 2022). Retention because of not meeting the Hungarian requirements is similarly stigmatized as retention because of bad performance in the Hungarian schools, so in the end these children also face a negative shock they have to cope with. Migration is most pronounced among families with higher educated parents who speak foreign languages well, although among the poorer, Roma families, also many migrate or flee to other countries in the hope for better living conditions (Árendás et al., 2022).

3.3 Descriptive statistics

3.3.1 Aspirations

Table 3.1 shows the educational aspirations in 6th grade by parental education (socioeconomic status). The sample of the table includes children who did not have to repeat any

[23] See `https://net.jogtar.hu/jogszabaly?docid=a1100190.tv`

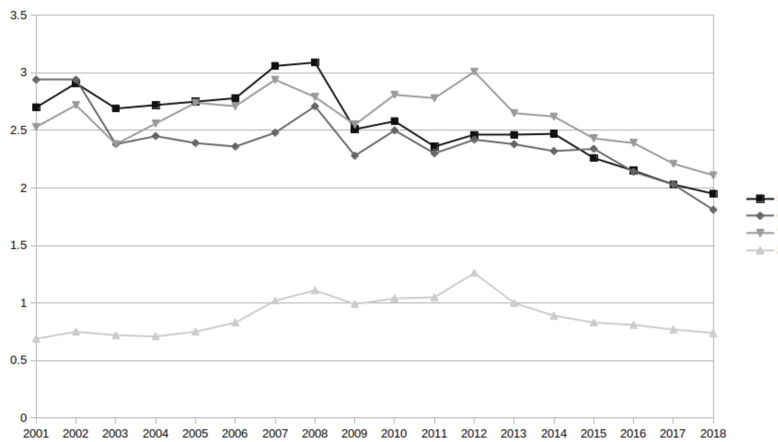


Figure 3.1: Percentage of retained students in higher grades of primary school

Source: KIR-STAT

grade until the 6^{th} grade[24]. Low-SES children are significantly more likely to aspire for at most vocational education already in 6^{th} grade than high-SES children (24.78% compared to 3.8%). They are also significantly less likely to aspire for tertiary education than high-SES children (25.12% compared to 69%). While the highest share of low-SES children (50.10%) aspire for a high-school diploma or a post-secondary non-tertiary qualification, for high-SES children the most popular education level is tertiary.

	Socio-economic status	
Aspirations in 6th grade	Low	High
At most vocational	24.78	3.80
High school diploma or post-secondary non-tertiary	50.10	27.20
College or university	25.12	69.00

Table 3.1: Educational aspirations of 6^{th} grade children by socioeconomic status

Note: Column percentages. Sample excludes children who have already repeated until 6th grade.

Figure 3.2 shows how educational aspirations change from 6^{th} to 8^{th} grade by SES and repeater status. I converted aspirations into into years of education for easier interpretation. On average, all non-repeaters increase their aspirations by a little (see Appendix Table C.2 for more detail). However, there are large differences between repeaters, especially at higher initial aspirations. For low-SES children, the aspiration gap between repeaters and non-repeaters increases with initial aspirations, while for high-SES children it stays pretty constant, at around a year. Appendix Table C.1 shows the number of observations and the share of repeaters in 7^{th} grade in each aspiration category– SES cell.

[24]In fact, low-SES children already have a much higher chance to repeat in the lower grades of primary school than high-SES children (7.5 percent vs. 1.5 percent in my sample). This means that the low-SES children who are selected into my sample were better able to meet school requirements

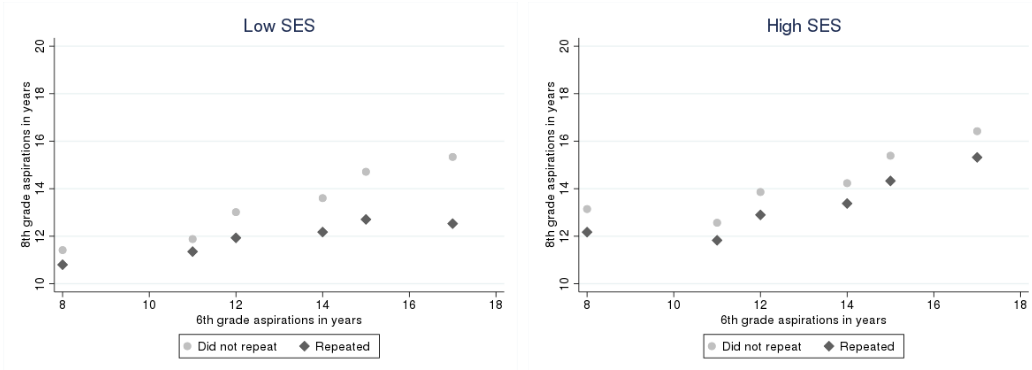


Figure 3.2: Change in educational aspirations between 6th and 8th grade by SES and repeating

Note: Educational aspirations are presented in years. At the upper end of the aspiration distribution, a master's degree and a doctoral degree are grouped into 17 years of education. The sample of the figures exclude children who were in an academic secondary school in 6th grade. It also excludes children who have already repeated until 7^{th} grade.

3.3.2 Retention

Table 3.2 shows the probability of children having to repeat 7^{th} grade by socioeconomic stastatustus, conditional on not having repeated until 7^{th} grade. Low-SES children are more likely to get retained: 1.24 percent of them repeat 7^{th} grade, compared to 0.33 percent of high-SES children. When we look at their 6^{th} year midterm grades in a core subject – mathematics – in Figure 3.3, we see that except for the lowest mark, low-SES children have a higher chance to repeat in each category. Among those who failed in mathematics in the first term of 6^{th} grade, high-SES are more likely to repeat 7^{th} grade than low-SES. This might be partly because low-SES in this group are more likely to get retained *already in* 6^{th} *grade*, so excluding 6^{th} grade repeaters introduces a selection into the sample. I will talk about this selection problem and potential solutions later in this section and in Section 3.7.

	Socio-economic status	
Grade retention in 7th grade	Low	High
Did not repeat	98.76	99.67
Repeated	1.24	0.33

Table 3.2: Repeating 7th grade grade by socioeconomic status

Note: Column percentages. The sample excludes children who have already repeated until 7th grade.

Since students from different socioeconomic backgrounds have a different chance to repeat 7^{th} grade at similar 6^{th} grade school performance, it is worth looking at repeaters

and catch up with their peers.

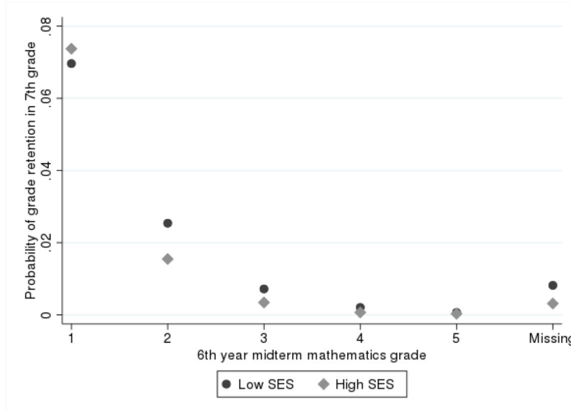


Figure 3.3: Probability of repeating 7th grade by 6th year midterm mathematics grade

Note: The sample excludes children who have already repeated until 7th grade.

and non-repeaters separately by SES to find out who the repeaters are in each group and what are the predictors of their retention. To be able to compare repeaters and non-repeaters by a wider range of characteristics, for this comparison I use the administrative database. Table 3.3 is a summary table of the characteristics of repeaters and non-repeaters by SES. Since the size of the administrative database is only half of the National ABC, I have around 500 low-SES repeaters and a bit over 200 high-SES repeaters in the sample, compared to the 55000 low-SES non-repeaters and 112000 high-SES non-repeaters. The variables I am comparing high-and low-SES children are factors that can affect retention: 6^{th} grade test scores and midterm grades in different subjects, whether their parents lived together in 6^{th} grade and whether they potentially separated between 6^{th} and 8^{th} grades (measured by the child living with both parents in 6^{th} grade but living with only one of them in 8^{th} grade), parental labor market status and its changes between 6^{th} and 8^{th} grades, visits to the general practitioner, and days spent in hospital. In general, both low- and high-SES repeaters perform worse in 6^{th} grade than non-repeaters, however, the average performance of high-SES children is better in all subjects and in both standardized tests compared to low-SES children. There are some differences regarding the family structure: high-SES repeaters are much more likely have parents not living together than low-SES repeaters, suggesting that being raised by a single parent might put a higher risk of retention on them. The share of separated parents are higher among repeaters in both SES groups. Another factor that might be a more frequent cause for retention for high-SES is being hospitalized: high-SES repeaters spend on average twice as much in hospital than low-SES repeaters (1.54 vs 0.75 days), while the difference is quite small for non-repeaters (0.24 vs 0.28 days).

However, when we control for these factors when regressing 7^{th} grade retention on socioeconomic status, high-SES children still seem to have a lower chance to repeat than low-SES. The first column of Table 3.4 shows the raw difference between the repeating probability of high- and low-SES children. Including the characteristics studied above, the gap between high- and low-SES children decreases significantly, but the difference is

	Low - Did not repeat	Low - Repeated	High - Did not repeat	High - Repeated
6th grade mathematics test	1441.95	1318.58	1563.89	1402.74
score	(173.12)	(146.07)	(174.06)	(167.27)
6th grade reading test score	1430.68	1296.33	1562.56	1379.33
	(174.38)	(148.46)	(174.04)	(179.21)
6th year midterm mathematics	3.19	2.06	3.91	2.36
grade	(1.02)	(0.73)	(0.96)	(0.99)
6th year midterm literature	3.63	2.48	4.28	2.88
grade	(0.98)	(0.81)	(0.82)	(1.04)
6th year midterm Hungarian	3.43	2.39	4.06	2.84
grammar grade	(0.97)	(0.77)	(0.88)	(1.01)
Parents lived together in 6th	0.77	0.67	0.79	0.50
grade	(0.42)	(0.47)	(0.41)	(0.50)
Parents separated from 6th to	0.05	0.11	0.04	0.12
8th grade	(0.22)	(0.31)	(0.21)	(0.33)
Mother does not work in 6th	0.37	0.47	0.15	0.23
grade	(0.48)	(0.50)	(0.36)	(0.42)
Mother has a permanent job in	0.50	0.40	0.74	0.62
6th grade	(0.50)	(0.49)	(0.44)	(0.49)
Father does not work in 6th	0.15	0.24	0.06	0.12
grade	(0.36)	(0.43)	(0.24)	(0.33)
Father has a permanent job in	0.64	0.54	0.72	0.65
6th grade	(0.48)	(0.50)	(0.45)	(0.48)
Mother stopped working between	0.07	0.09	0.04	0.09
6th and 8th grades	(0.26)	(0.29)	(0.19)	(0.28)
Father stopped working between	0.06	0.08	0.03	0.05
6th and 8th grades	(0.23)	(0.26)	(0.16)	(0.21)
Number of GP visits in 7th	4.99	7.57	4.05	6.31
grade	(4.82)	(6.74)	(4.07)	(6.87)
Days spent in hospital in 7th	0.28	0.75	0.24	1.54
grade	(1.76)	(4.82)	(1.70)	(9.02)
Observations	54812	493	112171	216

Table 3.3: Summary statistics of low- and high-SES 7th grade repeaters

Note: Rows 6-13 show shares of students with that characteristic, the rest of the rows show average levels. A parent not working means he/she is on childcare allowance, unemployed, retired, permanently ill/disabled or does not work for another reason. A parent having stopped working means he/she moved from working in any type of job to not working because of either of the reasons above.

still there, especially when including school fixed effects. Within the same school, high-SES children are less likely to get retained, even when controlling for a rich set of characteristics that affect retention. Column 4, however, shows that at the lowest 6^{th} year midterm mathematics grade, high-SES are equally likely to get retained than low-SES. Table 3.5 shows the predictors of repeating across 6^{th} year midterm grade groups. As the first column shows, SES does not affect retention in the lowest mathematics grade group, but it does affect retention at higher mathematics grades. For the lowest-performing students in mathematics, marks in other subjects and test scores do not seem to affect retention, neither family structure nor parental labor market status variables, or health characteristics. The only factor that still affects retention is the 6^{th} year behavior grade of the child, which is a measure of how well the child behaves at school according to her teachers. In this group, retention is 6.5 percent. The second and third column shows those children who got a 2 and at least a 3 in mathematics at 6^{th} grade midterm. The share of retained are 1.9 and 0.18 percent in these groups, and here, high-SES are ceteris paribus still less likely to repeat. For these groups, bad marks in another core subject, Hungarian literature, increase the risk of retention, as well as low effort and behavior grades, being raised by a single parent, parents separating at the time, and illnesses and hospitalization. These groups are therefore more heterogeneous in terms of the reason for repeating.

	Repeating 7th grade			
	(1)	(2)	(3)	(4)
High SES	-0.00783***	-0.000679*	-0.00149***	0.0103
	(0.00039)	(0.00037)	(0.00040)	(0.0099)
6th year midterm mathematics grade, baseline: 1 (fail)				
2		-0.0248***	-0.0247***	-0.0195***
		(0.0049)	(0.0049)	(0.0058)
3		-0.0304***	-0.0302***	-0.0276***
		(0.0048)	(0.0048)	(0.0057)
4		-0.0305***	-0.0301***	-0.0278***
		(0.0048)	(0.0049)	(0.0057)
5		-0.0294***	-0.0287***	-0.0263***
		(0.0048)	(0.0049)	(0.0057)
High SES × 2				-0.0170*
				(0.0100)
High SES × 3				-0.0109
				(0.0099)
High SES × 4				-0.0106
				(0.0099)
High SES × 5				-0.0107
				(0.0099)
Constant	0.0104***	0.0942***	0.0987***	0.0950***
	(0.00039)	(0.023)	(0.023)	(0.023)
Observations	243989	216529	216529	216529
Controls	no	yes	yes	yes
Year fixed effect	no	yes	yes	yes
School fixed effect	no	no	yes	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Regression of retention on SES and various characteristics.

Note: The controls include variables presented in Table 3.3. For all variables, dummies are used for missing values.

In the empirical analysis, first I look at the compensatory advantage of high socioeconomic status on aspirations on the whole sample, then I divide the sample by 6^{th} year mathematics grades to see if the compensatory advantage is different across these groups. Children with the lowest mark in mathematics most likely repeat because they perform badly at school, but the other two groups are more heterogeneous and have to be treated with caution. Mathematics is a fairly objective subject where students' knowledge can be easily assessed. However, other subjects give more leeway to the subjective judgment of teachers, and the advantages of higher socioeconomic background might show in other ways in these subjects: e.g. with similar lexical knowledge high-SES children might still be better in essay writing. These reasons would lead to high-SES students being retained in 7^{th} grade less likely than low-SES students. It is also possible that after having bad marks in 6^{th} grade but not having to repeat in the end, high-SES students start to catch up faster than low-SES students, so the lower chance of high-SES repeating 7^{th} grade might already capture some compensatory advantage in catching up after a risk of retention. Therefore, if I find a compensatory advantage of high SES in aspirations after retention, it is likely a lower bound of what I would find for the first retention shock and if high- and low-SES children of similar observable characteristics had equal chances to repeat. In the 3-5 math-

	Repeated 7th grade		
	(1)	(2)	(3)
	1	2	3-5
High SES	0.00432	-0.00888***	-0.000557*
	(0.018)	(0.0018)	(0.00029)
6th grade mathematics test score	0.0000177	-0.0000213***	-0.00000279***
	(0.000055)	(0.0000070)	(0.00000088)
6th grade reading test score	-0.000100	-0.0000236***	-0.000000833
	(0.000067)	(0.0000067)	(0.0000010)
6th year midterm literature grade	-0.0160	-0.00484***	-0.00120***
	(0.011)	(0.0013)	(0.00026)
6th year midterm Hungarian grammar grade	0.000548	-0.00209	0.000329
	(0.011)	(0.0013)	(0.00021)
6th year midterm effort grade	-0.00467	-0.00925***	-0.00101***
	(0.012)	(0.0015)	(0.00030)
6th year midterm behavior grade	-0.0281***	-0.00928***	-0.00153***
	(0.0093)	(0.0012)	(0.00026)
Parents lived together in 6th grade=1	-0.0195	-0.00813***	-0.000744**
	(0.016)	(0.0019)	(0.00029)
Parents separated from 6th to 8th grade=1	0.0515	0.0170***	0.00143**
	(0.035)	(0.0045)	(0.00066)
Mother does not work in 6th grade=1	-0.00340	-0.000228	0.000618
	(0.022)	(0.0027)	(0.00043)
Mother has a permanent job in 6th grade=1	-0.0136	0.000231	-0.000165
	(0.024)	(0.0027)	(0.00033)
Father does not work in 6th grade=1	0.0377	0.00388	0.000636
	(0.024)	(0.0031)	(0.00053)
Father has a permanent job in 6th grade=1	0.0108	-0.000320	-0.000213
	(0.020)	(0.0021)	(0.00025)
Mother stopped working between 6th and 8th grades=1	-0.00644	0.00578	0.00112
	(0.030)	(0.0037)	(0.00069)
Father stopped working between 6th and 8th grades=1	-0.0106	-0.00619*	0.000469
	(0.035)	(0.0035)	(0.00076)
Number of GP visits in 7th grade	0.00249	0.00160***	0.000175***
	(0.0016)	(0.00025)	(0.000046)
Days spent in hospital in 7th grade	-0.00275	0.00220**	0.00103***
	(0.0025)	(0.0011)	(0.00038)
Constant	0.300***	0.163***	0.0223***
	(0.10)	(0.013)	(0.0019)
Observations	3303	33271	180700
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Regression of retention on SES and various characteristics by 6th year midterm mathematics grade.

Note: The controls include variables presented in Table 3.3. For all variables, dummies are used for missing values. Mathematics grades 3-5 are pooled together because of a low number of repeaters in these groups.

ematics grade group the advantage of high-SES in retention is very low and significant only at the 10 percent level. The factors that affect the group who had a 2 in mathematics affect this group's retention, as well, but retention in this group is extremely rare.

3.3.3 8^{th} grade aspirations and secondary school tracks

Though unfortunately I cannot follow the students for long in the database, so I cannot look at how 8^{th} grade aspirations relate to later life outcomes, Figure 3.4 shows that they are good predictors of the type of secondary education students attend in 10^{th} grade. 66

10.14754/CEU.2022.08

percent of students with tertiary aspirations are in an academic secondary school at 10^{th} grade, and around 32 percent of them in a technical secondary school. Only 1.3 percent ends up in a vocational school. At the other end, 90 percent of students aspiring for at most a vocational certificate go to a vocational secondary school, 8 percent of them to a technical secondary, and around 2 percent to an academic secondary school. We have to note that students take the National ABC at the end of May, while 8^{th} graders are notified about their admission to secondary schools at the end of April, so when stating their aspirations most of them already know which secondary schools they will attend. Because of this we cannot treat 8^{th} grade aspirations and 10^{th} grade secondary school outcomes separately, as they are highly correlated. However, the high correlation also validates the aspirations measure, strengthening the interpretation that aspirations reflect children's true preferences about their future education.

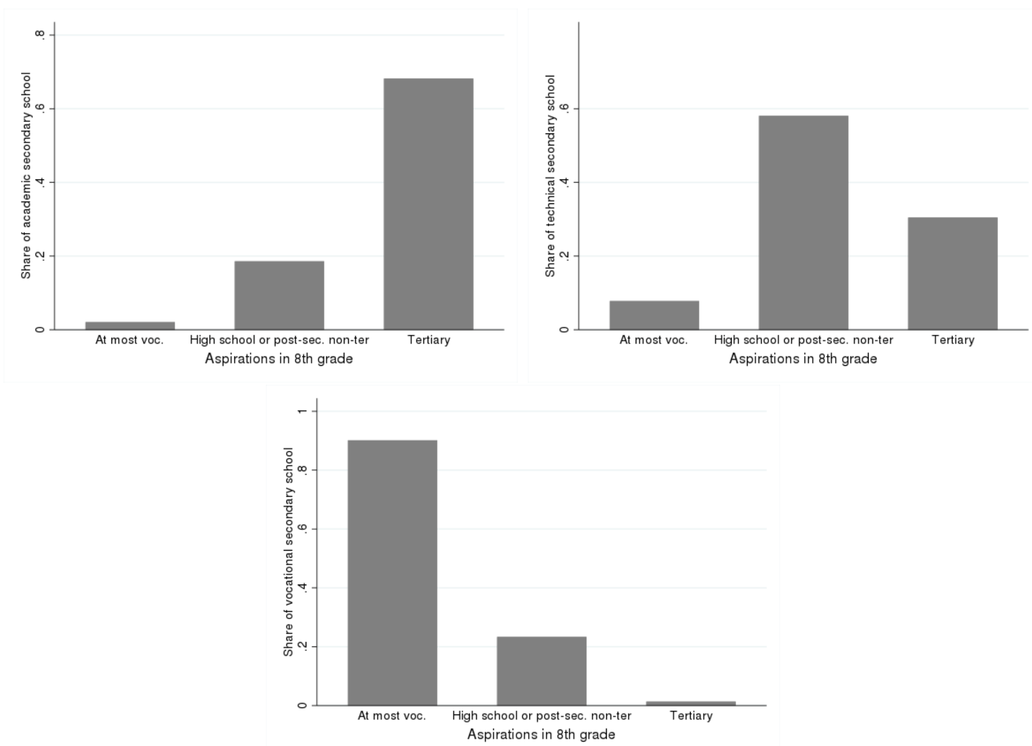


Figure 3.4: Probability of attending a type of secondary school by 8^{th} grade aspirations

Note: Percent of students with a particular 8th grade aspiration in different secondary education tracks in 10th grade.

3.3.4 Missing data

As completing the student background questionnaire is voluntary, there is a selection bias in my sample if completion is nonrandom. In the analysis I include observations with missing values by using dummies, however, I cannot apply this technique for socioeconomic status and the outcome variables. 6 percent of the sample has missing parental education data, so these observations have to be excluded from the analyses. Looking at 6^{th} grade mathematics test scores, the average in this sample with missing parental education is 1447 points, compared to the 1407 of low-SES and 1546 of high-SES children. In this sample the share of children I classify as repeaters is 5.2 percent, while it is 1.9 in the rest of the sample. Only 20 percent of this sample with missing data has data on 8^{th} grade educational aspirations: on average these children aspire for 13.7 years of education, compared to 13.3 years for low-SES and 15.4 for high-SES children. Based on these, there is a selection bias in completion towards children who perform better at school, have higher aspirations, and who most likely have higher educated parents. This means that the results I find might underestimate the compensatory advantage of high socioeconomic status, as I lose many observations of low-SES children.

3.4 Baseline empirical analysis

3.4.1 8^{th} grade aspirations

First I look at how grade retention in 7^{th} grade relates to 8^{th} grade aspirations and type of secondary education by socioeconomic status. In my main specifications, I estimate the following equation:

$$\begin{aligned} y_{i,s,t} = \alpha_0 &+ \alpha_1 \text{Repeated}_i + \alpha_2 \text{High SES}_i + \alpha_3 \text{Repeated}_i \times \text{High SES}_i \\ &+ \beta_1 X_i + \beta_2 X_{6,i,t} + \gamma_s + \eta_t + \epsilon_{i,s,t} \end{aligned} \tag{3.1}$$

Here, $y_{i,s,t}$ is one of two binary outcomes: one is 1 if student i from school s who was in 6^{th} grade in year t wants to reach at most vocational education in 8^{th} grade, and 0 if she has higher aspirations. The second outcome is 1 if the student wants to obtain a tertiary certificate in 8^{th} grade and 0 if she has lower aspirations. The main coefficients of interest are α_1 and α_3: α_1 shows how the aspirations of a low-SES child change if she has to repeat 7^{th} grade, and α_3 shows if this change is different for high-SES children. X_i are time-invariant characteristics of the child, such as gender and whether she has special education needs, and $X_{6,i,t}$ are other pre-treatment, 6^{th} grade characteristics, like 6^{th} grade GPA, mathematics and reading test scores, and the parents' labor market status. γ_s are 6^{th} grade school fixed effects, and η_t are 6^{th} grade year fixed effects.

Table 3.6 shows the effect of grade retention on the probability of aspiring for at most vocational education in 8^{th} grade. Column 1 shows the raw difference between repeaters and non-repeaters. Repeaters are 38.7 percentage points more likely to aspire for vocational education in 8^{th} grade than non-repeaters. In column 2 I include socioeconomic status and its interaction with repeating. We can see that low-SES repeaters are 37.8 percentage points more likely to aspire for vocational education than non-repeaters, while high-SES repeaters 21.1 percentage points. The differences remain large even if we control for 6^{th} grade aspirations and other time-invariant and 6^{th} grade characteristics in columns 3 and 4, and include school fixed effects in column 5. When including all controls and school fixed effects, low-SES repeaters are 15.6 percentage points more likely to aspire for vocational education than high-SES repeaters do not seem to change their vocational aspirations significantly.

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.378***	0.254***	0.153***	0.156***
	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)
High SES		-0.194***	-0.107***	-0.0821***	-0.0669***
		(0.0024)	(0.0016)	(0.0013)	(0.0013)
Repeated × High SES		-0.167***	-0.140***	-0.138***	-0.139***
		(0.021)	(0.019)	(0.018)	(0.018)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.376***	-0.289***	-0.275***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.431***	-0.294***	-0.273***
			(0.0033)	(0.0031)	(0.0030)
Constant	0.0991***	0.223***	0.515***	0.878***	0.813***
	(0.0019)	(0.0026)	(0.0036)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

Table 3.7 shows the effect of retention in 7^{th} grade on the probability of tertiary aspirations in 8^{th} grade. Repeaters are on average 42.8 percentage points less likely to aspire for tertiary education in 8^{th} grade than non-repeaters. This difference is heterogeneous by SES: in column 2 we can see that low-SES repeaters are 21.5 percentage points less likely to aspire for tertiary education than non-repeaters, but for high-SES repeaters the difference is even higher: -42.9. Although the difference is much higher for high-SES children, still 27.2 percent of repeaters in this group aspire for tertiary education, while for low-SES children this share drops from 25.4 percent to 3.9 percent. When controlling for 6^{th} grade aspirations (column 3) other 6^{th} grade controls (column 4), and school fixed effects (column 5), the difference between low-SES repeaters and non-repeaters decreases sharply, and disappears. For high-SES repeaters the difference remains: they are around 7.9 percentage points less likely to aspire for tertiary education than non-repeaters.

Table 3.8 explores Figure 3.2 in a regression framework, namely, whether the aspirations of repeaters and non-repeaters change differently by how high their initial aspirations were.

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.215***	-0.0718***	0.00490	0.00200
	(0.0072)	(0.0053)	(0.0058)	(0.0056)	(0.0056)
High SES		0.447***	0.229***	0.177***	0.147***
		(0.0032)	(0.0024)	(0.0022)	(0.0018)
Repeated × High SES		-0.214***	-0.133***	-0.0849***	-0.0789***
		(0.018)	(0.016)	(0.015)	(0.015)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.141***	0.0440***	0.0365***
post-secondary non-tertiary			(0.0019)	(0.0019)	(0.0019)
College or university			0.646***	0.400***	0.373***
			(0.0023)	(0.0026)	(0.0026)
Constant	0.541***	0.254***	0.00958***	-0.536***	-0.425***
	(0.0043)	(0.0021)	(0.0020)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Probability of 8^{th} grade aspirations being tertiary education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

Here I run the following regression:

$$\begin{aligned} y_{i,s,t} = \alpha_0 &+ \alpha_1 \text{Repeated}_i + \alpha_2 \text{High SES}_i + \alpha_3 \text{Repeated}_i \times \text{High SES}_i \\ &+ \alpha_4 \text{6th grade aspirations}_i + \alpha_5 \text{Repeated}_i \times \text{6th grade aspirations}_i \\ &+ \alpha_6 \text{6th grade aspirations}_i \times \text{High SES}_i \\ &+ \alpha_7 \text{Repeated}_i \times \text{6th grade aspirations}_i \times \text{High SES}_i + \beta_1 X_i + \beta_2 X_{6,i,t} \\ &+ \gamma_s + \eta_t + \epsilon_{i,s,t} \end{aligned} \tag{3.2}$$

6^{th} grade aspirations are now measured by the years of education the child wants to achieve. α_5 shows whether children with initially high aspirations degrade their aspirations more after grade retention, and α_7 shows if this decrease is heterogeneous by socioeconomic status. 6^{th} grade aspirations are demeaned, so the baseline of the interactions always show the effects at the average level of initial aspirations. As we can see in column 1 of Table C.9, non-repeaters aspire for almost 15 years of education on average, which corresponds to a college or BA-level degree. Repeaters aspire for 2.5 years less, which is nearly equivalent to a high school diploma. When we control for socioeconomic status in column 2, we see that high-SES non-repeaters aspire for almost 2 years more education than low-SES non-repeaters, but the drop after retention in aspirations is higher for high-SES than for low-

SES, by a third of a year. The negative coefficient on the interaction can be explained by high-SES children having initially higher aspirations, so they can decrease more on average with retention. Controlling for 6^{th} grade aspirations, the difference between repeaters and non-repeaters by parental education virtually disappears. On average, there is no compensatory advantage of high SES in aspirations. However, when in column 4 I interact retention and parental education with 6^{th} grade aspirations, the average negative effect seems to mask a heterogeneous effect by initial aspirations. Column 5 adds 6^{th} grade controls, while column 6 adds school fixed effects. Column 6 shows that a low-SES child with average 6^{th} grade aspirations (14.5 years) decreases her aspirations by 10 months (0.85 years) in case of having to repeat 7^{th} grade. In contrast, high-SES repeaters with average 6^{th} grade aspirations decrease their aspirations by only 3 months. Retention causes a larger drop the higher the initial aspirations were for low-SES children, by about 2 months (0.163 years) for each year of 6^{th} grade aspirations. However, having high socioeconomic status completely offsets this penalty of higher initial aspirations.

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.462***	-1.734***	-0.943***	-1.691***	-0.832***	-0.850***
	(0.043)	(0.043)	(0.042)	(0.078)	(0.072)	(0.070)
High SES		1.883***	0.827***	0.840***	0.559***	0.446***
		(0.016)	(0.0094)	(0.0094)	(0.0075)	(0.0064)
Repeated ×		-0.345***	-0.0184	0.687***	0.601***	0.599***
High SES		(0.097)	(0.082)	(0.12)	(0.11)	(0.11)
6th grade			0.555***	0.525***	0.341***	0.320***
aspirations in years			(0.0019)	(0.0027)	(0.0027)	(0.0026)
Repeated ×				-0.279***	-0.163***	-0.163***
6th grade aspirations in years				(0.023)	(0.021)	(0.021)
High SES ×				0.0549***	0.0328***	0.0193***
6th grade aspirations in years				(0.0033)	(0.0030)	(0.0029)
Repeated ×				0.202***	0.139***	0.145***
High SES × 6th grade aspirations in years				(0.042)	(0.039)	(0.038)
Constant	14.71***	13.50***	14.12***	14.08***	11.06***	11.34***
	(0.019)	(0.013)	(0.0079)	(0.0085)	(0.16)	(0.15)
Observations	391412	391412	391412	391412	391412	391412
6th grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the class level. Clustered standard errors in parentheses.

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

Table 3.8: Educational aspirations in 8^{th} grade in years

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma and high SES means at least one parent has a high school diploma. 6th grade aspirations are demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

3.5 Subsamples of repeaters by 6^{th} grade mathematics performance

To be able to separate children with different reasons for repeating, now I present the results of the regressions similar to Equation 3.2 ran on subsamples based on 6^{th} grade midterm mathematics performance. For these regressions, I also use the National ABC database, because there is already half of the sample in the Admin3 dataset, and with the very few repeaters I would have too few observations in each of these subsamples. The drawback of using the National ABC is that I cannot control for GP visits and hospitalizations, which, as we saw in Table 3.5, affect retention in the groups with better 6^{th} grade mathematics performance. Appendix Section C.0.3.4 presents the regressions below ran on the Admin3 data and including controls for GP visits and days spent in a hospital.

Table 3.9 shows the results for the lowest performer students in mathematics. The share of repeaters is 7.1 percent in this group, and it is quite similar by SES: 6.9 percent of low-, and 7.4 percent of high-SES children in this group repeat. The first column of Table 3.9 shows that within this group, repeating does not decrease future aspirations significantly for either low- or high-SES children. High-SES non-repeaters, on the other hand, aspire for about 8.5 months (0.7 years) higher education than low-SES non-repeaters. The second column controls for initial aspirations, and the third one interacts aspirations with socioeconomic status, assuming that socioeconomic status has a differential compensating effect at different levels of aspirations. When conrolling for initial aspirations, we see that in this group, repeaters do decrease their – already low, about 12 years – aspirations but only by around 3 months (0.24 years), and the decrease is only significant on the 10 percent level.

The signs of the coefficients, except for the triple interaction of repeating, high SES, and 6^{th} grade aspirations, are similar to the full-sample ones, though they are smaller in magnitude. Part of the reason for the smaller magnitude is that children with bad mathematics performance already had low aspirations in 6^{th} grade. Combined with the low sample size of this group,[25] the interactions of repeating and SES are not significant. However, the coefficient on the triple interaction is negative, meaning that high-SES children with higher initial aspirations in this group decrease their aspirations more than low-SES children if they repeat. Although the coefficient is insignificant, the opposite sign suggests that in this low-performer group, both low- and high-SES children suffer from retention. socioeconomic status does seem to matter though for the non-repeaters. Low-performer children who in the end did not have to repeat the 7^{th} grade aspire for higher education levels if they are from high SES, and the difference is even higher for children with higher initial aspirations.

Appendix Table C.11 shows the results of the same regression ran on the Admin3 data with controls for GP visits and hospitalization in 7^{th} grade as well. The only coefficients that are large and significant are socioeconomic status for non-repeaters – they aspire for 7 months (0.6 years) higher education than low-SES –, and 6^{th} grade aspirations. Low-SES repeaters decrease their aspirations by 3 months (0.24 years), while high-SES by 3.5 months (0.3 years), though none of the differences are significant. This sample is very small, so

[25] The number of repeaters are 281 in the low-SES, and 114 in the high-SES group.

the non-significant results might be because of low sample size.[26] Each extra year of 6^{th} grade aspirations decreases repeaters' 8^{th} grade aspirations, and for high-SES repeaters the decrease is even larger, though these differences are not significant, either. It seems that in this low-performer group, compensatory advantage does not work after retention, leaving both high- and low-SES children with lowered expectations about their educational attainment.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.198	-0.222*	-0.239*
	(0.14)	(0.13)	(0.13)
High SES	0.695***	0.515***	0.460***
	(0.069)	(0.067)	(0.067)
Repeated ×	-0.0719	-0.0193	0.0945
High SES	(0.25)	(0.24)	(0.25)
6th grade		0.254***	0.228***
aspirations in years		(0.017)	(0.020)
Repeated ×			-0.0385
6th grade aspirations in years			(0.066)
High SES ×			0.111***
6th grade aspirations in years			(0.034)
Repeated ×			-0.122
High SES × 6th grade aspirations in years			(0.12)
Observations	6338	6338	6338
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

Table 3.10 shows the results on the group who have just passed mathematics at 6^{th} grade midterm. In this group, 2.2 percent had to repeat 7^{th} grade (2.5 percent of low-SES, and 1.5 of high-SES).[27] Children in this group aspire for 13 years of education in 6^{th} grade on average. Repeaters have around half a year lower aspirations in 8^{th} grade than non-repeaters, and without controlling for aspirations, this difference is similar for high- and low-SES children. The second column controls for 6^{th} grade aspirations, and the third column introduces the triple interaction. In the third column, we see similar patterns to the full sample: there is a compensatory advantage of high socioeconomic status in aspirations for repeaters with average initial aspirations. While low-SES repeaters decrease their aspirations by half a year, for high-SES repeaters this decrease is only a fifth of a year. Each extra year of 6^{th} grade aspirations increases the drop in aspirations after retention

[26] In the estimation sample there are 117 low-SES repeaters and 55 high-SES repeaters.
[27] This gives 815 low-SES and 301 high-SES repeaters.

by a little bit more than a month. This increase is lower for high-SES children, though not significantly. The differences in the aspirations of high- and low-SES non-repeaters are there in this group, too.

Appendix Table C.12 presents the results of the same regression on the Admin3 data, including controls for 7^{th} grade health characteristics.[28] The picture here is similar to that in Table 3.10, except that while low-SES repeaters with average 6^{th} grade aspirations decrease their 8^{th} grade aspirations by about half a year, high-SES repeaters do not decrease them significantly. While the repeater–non-repeater aspiration gap increases for low-SES children by 1 month per each year of initial aspirations, the gap is constant and not significantly different from zero for high-SES children. In this group there is already significant compensatory advantage of high socioeconomic status.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.474***	-0.431***	-0.542***
	(0.050)	(0.050)	(0.061)
High SES	0.723***	0.519***	0.504***
	(0.016)	(0.015)	(0.015)
Repeated ×	0.100	0.197*	0.335***
High SES	(0.12)	(0.11)	(0.12)
6th grade		0.283***	0.264***
aspirations in years		(0.0044)	(0.0053)
Repeated ×			-0.107***
6th grade aspirations in years			(0.028)
High SES ×			0.0554***
6th grade aspirations in years			(0.0078)
Repeated ×			0.0694
High SES × 6th grade aspirations in years			(0.060)
Observations	61465	61465	61465
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: 8^{th} grade aspirations for those only passing mathematics at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

Finally, Table 3.11 shows the effect of retention on aspirations for those whose 6^{th} grade midterm mathematics performance was at least average. In this group, retention was extremely rare: 0.23 percent on average, with 0.45 for low- and 0.13 for high-SES students.[29] The average of 6^{th} grade aspirations is 15 years in this group. On average, there does not seem to be a compensatory advantage for repeaters in aspirations in this

[28] In this sample, there are 380 low-SES repeaters and 118 high-SES repeaters.

[29] 311 low-SES and 218 high-SES repeaters.

group either. However, in the third column, where we assume that SES helps children differently at different initial aspirations, there is a large compensatory advantage at the average aspirations, that increases even more if initial aspirations are higher. Low-SES repeaters decrease their aspirations by a whole year in this group, while for high-SES repeaters the decrease is only 4.3 months (0.36 years). Every extra year of 6^{th} grade aspirations increase the repeater–non-repeater gap by 2 months (0.16 years) for low-SES, but for high-SES, the gap is constant.

Appendix Table C.13 controls for GP visits and hospital stays in the Admin3 database. Controlling for these characteristics (and having half the sample size of the National ABC database), the compensatory advantage of high SES in this group becomes insignificant, though still large and positive. The number of repeaters in this sample is also quite small (147 low-SES repeaters and 110 high-SES repeaters), so it is possible that there is still a compensatory advantage among children repeating for similar reasons, but it is more heterogeneous than in the group of children just passing mathematics in 6^{th} grade.

It is likely, therefore, that the compensatory advantage I find on the overall sample is driven by repeaters in the higher performer groups, who are more likely repeating because of health shocks, or other, unobservable reasons. It is also possible that some children in this group repeat 7^{th} grade because the family spent a year abroad; a factor that I do not observe in the dataset. However, as Appendix Table C.4 shows, even in the highest performer group, repeaters' 6^{th} year test scores and grades are worse than non-repeaters', and they visited the GP more times and spent more days in a hospital than non-repeaters. Based on these it is unlikely that this group is mostly composed of children who spent a year abroad with their family. Even if there were many such children, retention because of not being able to meet the Hungarian requirements after returning from abroad is also stigmatized (Árendás et al., 2022), so it is not entirely wrong to treat this type of retention as a similar negative shock to other types of retention.

3.6 10^{th} grade outcomes

Finally I look at how the aspiration changes translate to changes in the probability of attending a secondary school that gives access to tertiary education, so one that ends with a high school diploma. In Hungary in my sample period these were the academic secondary schools and the technical secondary schools. I estimate the same equation as Equation 3.2, except that the outcome variable here is a dummy indicating whether the child is in a technical or an academic secondary school in 10^{th} grade or in a vocational school that does not provide access to tertiary education. I look at the results in the three subsamples that I analysed in the previous section: those who failed mathematics in 6^{th} grade midterm, those who only passed, and those who had higher marks. Tables 3.12-3.14 show the results of the regressions by subsample.

For the lowest performers in mathematics, the patterns are similar to the aspiration changes – partly because in 8^{th} grade, when stating their aspirations, children already know which secondary school they were admitted to. In Table 3.12 we see that repeaters are 11 percentage points less likely to attend an institution giving access to tertiary education than non-repeaters, and for each extra year of initial aspirations, the probability of attending this institution decreases by 5.9 percentage points. This decrease is not compensated

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.702***	-0.582***	-1.005***
	(0.089)	(0.088)	(0.16)
High SES	0.703***	0.452***	0.459***
	(0.0079)	(0.0070)	(0.0072)
Repeated ×	0.259	0.186	0.644***
High SES	(0.16)	(0.15)	(0.21)
6th grade		0.345***	0.337***
aspirations in years		(0.0021)	(0.0031)
Repeated ×			-0.164***
6th grade aspirations in years			(0.051)
High SES ×			0.0136***
6th grade aspirations in years			(0.0035)
Repeated ×			0.212***
High SES × 6th grade aspirations in years			(0.079)
Observations	309209	309209	309209
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11: 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

significantly by SES at either initial aspiration level. High-SES non-repeaters, on the other hand, are more likely to attend this institution at every aspiration level than low-SES non-repeaters. When using the Admin3 database and including GP visits and hospital days in Appendix Table C.14, the picture is similar, though because of the even lower sample size, none of the interactions with high SES are significant. Comparing children who repeated for the same reason, repeater low-SES are 19 percentage points less likely to attend a secondary school giving a high school diploma, and each extra initial year of aspirations widens this gap by 14 percentage points. The chances of high-SES repeaters at the average aspiration level to attend this institution are at least as much lower than low-SES repeaters' compared to non-repeaters, though initial aspirations seem to – though non-significantly – compensate for them. The advantage of high-SES non-repeaters remains significant.

Interestingly, while for children just passing we saw a large compensatory advantage for repeaters at the average aspirations, this compensatory advantage does not carry on strongly to an advantage in attending a secondary school ending with a high school diploma. Low-SES repeaters are 6 percentage points less likely to attend this institution than non-repeaters, though the difference is only significant on the 10 percent level. For high-SES repeaters the measure of the aspiration gap is quite noisy, but for them, the decrease in aspirations is not significantly different from zero. Table C.15 controls for GP visits and hospital days, as well. The aspiration gap for low-SES children remain similar, though it

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0846*	-0.0847	-0.108**
	(0.051)	(0.052)	(0.049)
High SES	0.182***	0.150***	0.141***
	(0.029)	(0.029)	(0.030)
Repeated ×	0.0208	0.0317	0.0588
High SES	(0.11)	(0.11)	(0.11)
6th grade		0.0414***	0.0355***
aspirations in years		(0.0070)	(0.0087)
Repeated ×			-0.0596***
6th grade aspirations in years			(0.022)
High SES ×			0.0199
6th grade aspirations in years			(0.014)
Repeated ×			0.0861
High SES × 6th grade aspirations in years			(0.053)
Observations	3338	3338	3338
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

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

Table 3.12: Attending a secondary school ending with a high school diploma (1 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

becomes insignificant. There seems to be no compensatory advantage of high SES when comparing children repeating for similar reasons; the coefficient on the interaction of high SES and repeating even becomes negative. The number of repeaters here is very small, as well. The advantage of high-SES non-repeaters is strong and significant in this sample, too.

There is a large compensatory advantage of high socioeconomic status in the sample of students performing well in mathematics. Here repeaters are, both statistically and economically significantly, less likely to attend an institution ending with a high school diploma than non-repeaters. Low-SES repeaters are 25 percentage points less likely to attend this institution than non-repeaters, while for high-SES children, the difference is non-significant. The compensatory advantage is even higher at higher initial aspirations. The higher 6^{th} grade aspirations are, the more likely low-SES non-repeaters are attending a secondary school that gives access to tertiary education, but for high-SES non-repeaters, the chances of attending this institution does not depend so strongly on initial aspirations.

Appendix Table C.16 includes GP visits and hospital days and uses the Admin3 database. The compensatory advantage found without the health controls disappears when we compare children who repeated for the same reason. Without controlling for the interactions of initial aspirations, repeaters are 22 percentage points less likely to attend a high

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0637**	-0.0519**	-0.0610*
	(0.026)	(0.025)	(0.033)
High SES	0.172***	0.136***	0.136***
	(0.0056)	(0.0055)	(0.0056)
Repeated ×	0.0451	0.0444	0.0516
High SES	(0.046)	(0.045)	(0.051)
6th grade		0.0531***	0.0536***
aspirations in years		(0.0015)	(0.0019)
Repeated ×			-0.00978
6th grade aspirations in years			(0.016)
High SES ×			-0.00115
6th grade aspirations in years			(0.0027)
Repeated ×			0.0163
High SES × 6th grade aspirations in years			(0.026)
Observations	37074	37074	37074
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

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

Table 3.13: Attending a secondary school ending with a high school diploma (2 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

school providing access to tertiary education, and this difference is similar for high- and low-SES children. However, when we control for the interactions of initial aspirations, the aspiration gap at the average initial aspiratoins becomes insignificant for both SES groups. Those low-SES repeaters who had higher initial aspirations are 2.5 percentage points more likely to attend this institution per each year of initial aspirations, while for high-SES the chances of attending does not depend on initial aspirations.

3.7 Discussion

I find that children from different socioeconomic backgrounds cope differently with having to repeat the 7^{th} grade of primary school: all repeaters decrease their aspirations, but the magnitudes are larger for low-SES children. The post-retention SES-gap in aspirations is the highest with initially high aspirations. When looking at subsamples by previous academic performance in a core subject, mathematics, the picture changes sharply. Students with poor prior mathematics performance decrease their aspirations if they get retained in 7^{th} grade, but the decrease is similar for all children, regardless of socioeconomic background. When controlling for factors that proxy missing school because of health issues, the coefficients on compensatory advantage even turn negative, though insignificant. These children are then all less likely to end up in a secondary school giving access to tertiary ed-

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.204***	-0.198***	-0.241***
	(0.047)	(0.047)	(0.072)
High SES	0.0841***	0.0654***	0.0440***
	(0.0018)	(0.0017)	(0.0016)
Repeated ×	0.148**	0.144**	0.194**
High SES	(0.061)	(0.060)	(0.080)
6th grade		0.0256***	0.0506***
aspirations in years		(0.00050)	(0.00087)
Repeated ×			-0.0293
6th grade aspirations in years			(0.023)
High SES ×			-0.0404***
6th grade aspirations in years			(0.00096)
Repeated ×			0.0774***
High SES × 6th grade aspirations in years			(0.029)
Observations	212904	212904	212904
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

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

Table 3.14: Attending a secondary school ending with a high school diploma (3-5 in mathematics)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, and whether they separated between 6th and 8th grades. For all variables, dummies are used for missing vwill lalues. 6th grade aspirations are demeaned.

ucation, but there is no difference by socioeconomic status here, either. However, in these groups, socioeconomic status has an advantage for the non-repeaters in both aspirations and secondary school track. It is possible therefore that, in line with the compensatory advantage literature, those students who were on the edge of repeating but in the end they did not have to, are able to catch up better at school if they have a better socioeconomic background. However, retention might be, in line with most findings in the retention literature, such a big negative shock that even high socioeconomic status cannot offset it.

On the other hand, I find a compensatory advantage for those children who performed better previously, in both 8^{th} grade aspirations and secondary track choice. Retention is quite rare in these groups, especially among those with good mathematics performance, and is likely affected by other factors – such as health issues, separation of parents, or spending a year abroad with the family – than only bad performance. Except for the lowest-performer students, high-SES children always have a lower chance to repeat conditional on all observable factors that can affect retention. It is likely that, since I excluded children who repeated the 6^{th} grade, I introduced an extra selection into the treatment: those children who almost had to repeat 6^{th} grade but they did not have to in the end, might differ by SES in terms of how they catch up with the school workload and how much effort they put into avoiding further risk of retention. The compensatory advantage I find in these groups might be a lower bound of the compensatory advantage I could estimate had I not introduced a selection by excluding repeaters in the 6^{th} grade, so if high-SES children had the same chance to repeat 7^{th} grade as low-SES. In these groups, when controlling for GP visits and hospital days, the compensatory advantage in aspirations remain but the compensatory advantage in attending a secondary school that provides access to tertiary education diminishes, though the sample sizes here are very small.

What could be the mechanisms behind the compensatory advantage? Families try to avoid downward mobility by parents pushing their children towards education levels at least as high as their own. After a negative shock, such as retention, parents try to still push their children towards better education, but high-SES parents have more resources to do so, and they also start from a higher reference point (Bernardi, 2014). Another reason why I see different responses in aspirations by high- and low-SES children might be that their parents are not equally involved in completing the questionnaire. Müller (2021) finds that the aspiration gap between high- and low-SES children is smaller if parents are not informed of children's answers to the aspiration question. The result is driven by high-SES children who aspire for tertiary education when their parents see their answers but for lower education when they do not. It is possible that, while in 6^{th} grade, both high- and low-SES parents help their children completing the questionnaire, in 8^{th} grade, children mostly do it on their own, but the parents of retained students still monitor their children's answers, leading to still high goals for high-SES, but reduced goals for low-SES children.

10.14754/CEU.2022.08

References

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden, "Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians?," *Journal of Political Economy*, 2020, (262675).

Almås, Ingvild, Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden, "Fairness and the Development of Inequality Acceptance," *Science*, 2010, *328* (5982), 1176–1178.

Andre, Peter, "Shallow Meritocracy: An Experiment on Fairness Views," *SSRN Electronic Journal*, 2021, *2021*, 1–79.

Árendás, Zsuzsanna, Judit Durst, Noémi Katona, and Vera Messing, "The Limits of Trading Cultural Capital: Returning Migrant Children and their Educational Trajectory in Hungary," in Adrienne Lee Atterberry, Derrace Garfield McCallum, Siqi Tu, and Amy Lutz, eds., *Children and Youths' Migration in a Global Landscape*, Emerald Publishing Limited, 2022, pp. 115–139.

Augenblick, Ned, "Short-Term Time Discounting of Unpleasant Tasks," July 2018.

_ **and Matthew Rabin**, "An experiment on time preference and misprediction in unpleasant tasks," *Review of Economic Studies*, 2019, *86* (3), 941–975.

_ **, Muriel Niederle, and Charles Sprenger**, " Working over Time: Dynamic Inconsistency in Real Effort Tasks *," *The Quarterly Journal of Economics*, 05 2015, *130* (3), 1067–1115.

Baucells, Manel and Franz H. Heukamp, "Probability and Time Trade-Off," *Management Science*, 2012, *58* (4), 831–842.

Benndorf, Volker, Holger A. Rau, and Christian Sölch, "Minimizing learning behavior in repeated real-effort tasks," *Center for European, Governance and Economic Development Research Discussion Papers*, 3 2018.

Bernardi, Fabrizio, "Compensatory Advantage as a Mechanism of Educational Inequality: A Regression Discontinuity Based on Month of Birth," *Sociology of Education*, 2014, *87* (2), 74–88.

_ **and Héctor Cebolla-Boado**, "Previous School Results and Social Background: Compensation and Imperfect Information in Educational Transitions," *European Sociological Review*, 10 2013, *30* (2), 207–217.

_ **and Michael Grätz**, "Making up for an unlucky month of birth in school : causal evidence on the compensatory advantage of family background in England," *Sociological Science*, 2015, *2*, 235–251.

Bradley, Robert, Robert Corwyn, Bettye Caldwell, Leanne Whiteside-Mansell, Gail Wasserman, and Iris Mink, "Measuring the Home Environments of Children in Early Adolescence," *Journal of Research on Adolescence - J RES ADOLESCE*, 07 2000, *10*, 247–288.

Cappelen, Alexander, Karl Ove Moene, Siv-Elisabeth Skjelbred, and Bertil Tungodden, "The Merit Primacy Effect," Working Papers 2017-047, Human Capital and Economic Opportunity Working Group 2017.

Cappelen, Alexander W., Erik Sørensen, and Bertil Tungodden, "Responsibility for what? Fairness and individual responsibility," *European Economic Review*, 2010, *54* (3), 429–441.

_ **, James Konow, Erik Ø. Sørensen, and Bertil Tungodden**, "Just Luck: An Experimental Study of Risk-Taking and Fairness," *American Economic Review*, June 2013, *103* (4), 1398–1413.

_ **, Thomas de Haan, and Bertil Tungodden**, "Fairness and limited information: Are people Bayesian meritocrats?," 2020. Presentation.

Chakraborty, Anujit, Yoram Halevy, and Kota Saito, "The Relation between Behavior under Risk and over Time," *American Economic Review: Insights*, March 2020, *2* (1), 1–16.

Cham, Heining, Jan N. Hughes, Stephen G. West, and Myung Hee Im, "Effect of retention in elementary grades on grade 9 motivation for educational attainment," *Journal of School Psychology*, 2015, *53* (1), 7–24.

Chen, Daniel L., Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88 – 97.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Cunningham, Tom and Jonathan de Quidt, "Implicit Preferences," *Available at SSRN 2709914*, 2022.

Deffains, Bruno, Romain Espinosa, and Christian Thöni, "Political self-serving bias and redistribution," *Journal of Public Economics*, 2016, *134*, 67–74.

Drucker, Luca Flóra, "Meritocracy and Persistent Luck," December 2021.

_ **and Marc Kaufmann**, "Excuse-driven Present Bias," *AEA RCT Registry*, September 2019.

Durante, Ruben, Louis Putterman, and J. van der Weele Joël, "Preferences for redistribution and perception of fairness: An experimental study," *Journal of the European Economic Association*, 2014, *12* (4), 1059–1086.

Eisenkopf, Gerald, Urs Fischbacher, and Franziska Föllmi-Heusi, "Unequal opportunities and distributive justice," *Journal of Economic Behavior and Organization*, 2013, *93*, 51–61.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *The Review of Economic Studies*, 2016, *83* (2), 587–628.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences*," *The Quarterly Journal of Economics*, 05 2018, *133* (4), 1645–1692.

_ , Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers, "Socio-Economic Status and Inequalities in Children's IQ and Economic Preferences," *Journal of Political Economy*, 2019.

Fedyk, Anastassia, "Asymmetric naivete: Beliefs about self-control," *Available at SSRN 2727499*, 2021.

Fehr, Dietmar and Martin Vollmann, "Misperceiving Economic Success : Experimental Evidence on Meritocratic Beliefs and Inequality Acceptance," 2020, (695).

Gary-Bobo, Robert J., Marion Goussé, and Jean-Marc Robin, "Grade retention and unobserved heterogeneity," *Quantitative Economics*, 2016, *7* (3), 781–820.

Genicot, Garance and Debraj Ray, "Aspirations and Inequality," *Econometrica*, 2017, *85* (2), 489–519.

_ and _ , "Aspirations and economic behavior," *Annual Review of Economics*, 2020, *12*, 715–746.

Heath, Chip, Richard P. Larrick, and George Wu, "Goals as Reference Points," *Cognitive Psychology*, 1999, *38*, 79–109.

Heckman, James J. and Stefano Mosso, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

Hughes, Jan N., Oi-Man Kowk, and Im MyungHee, "Effect of Retention in First Grade on Parents' Educational Expectations and Children's Academic Outcomes," *American Educational Research Journal*, 2013, *50* (6).

Jacob, Brian A. and Lars Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *The Review of Economics and Statistics*, February 2004, *86* (1), 226–244.

— **and** — , "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, 2009, *1* (3), 33–58.

Kahneman, Daniel and Amos Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, 1979, pp. 263–291.

Kőszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences*," *The Quarterly journal of economics*, 2006, *121* (4), 1133–1165.

Konow, James, "Fair Shares : Accountability and Cognitive Dissonance in Allocation Decisions," *The American Economic Review*, 2000, *90* (4), 1072–1091.

Mott, Frank L., "The Utility of the HOME-SF Scale for Child Development Research in a Large National Longitudinal Survey: The National Longitudinal Survey of Youth 1979 Cohort," *Parenting*, 2004, *4* (2-3), 259–270.

Müller, Maximilian, "Intergenerational Transmission of Education: Internalized Aspirations versus Parent Pressure," Working Paper 2021.

O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *The American Economic Review*, March 1999, *89* (1), 103–124.

OECD, *Education at a Glance 2015* 2015.

Page, Lionel, Louis Levy Garboua, and Claude Montmarquette, "Aspiration levels and educational choices: An experimental study," *Economics of Education Review*, 2007, *26* (6), 747–757.

Pekkarinen, Tuomas, "School tracking and intergenerational social mobility," *IZA World of Labor*, 2018, (56).

Rodriguez-Lara, Ismael and Luis Moreno-Garrido, "Self-interest and fairness: Self-serving choices of justice principles," *Experimental Economics*, 2012, *15* (1), 158–175.

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Wössmann, "Education policy and equality of opportunity," *Kyklos*, 2008, *61* (2), 279–308.

Sebők, Anna, "The Panel of Linked Administrative Data of CERS Databank," *Budapest Working Papers on the Labour Market*, 2019, (2).

Simon, Herbert A., "Designing organizations for an information-rich world," in M. Greenberger, ed., *Computers, Communications, and the Public Interest*, Johns Hopkins Press, 1971, pp. 37–72.

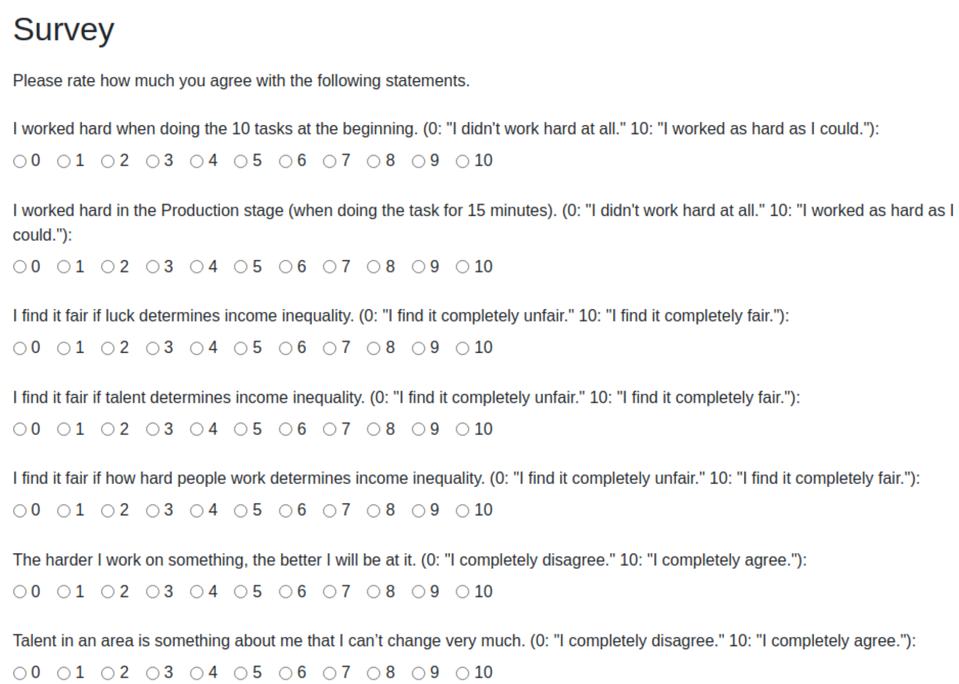
Tversky, Amos and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, 1981, *211* (4481), 453–458.

Yaouanq, Yves Le and Peter Schwardmann, "Learning about one's self," 2019.

10.14754/CEU.2022.08

A Appendix for Chapter 1

A.1 Design



Survey

Please rate how much you agree with the following statements.

I worked hard when doing the 10 tasks at the beginning. (0: "I didn't work hard at all." 10: "I worked as hard as I could."):

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

I worked hard in the Production stage (when doing the task for 15 minutes). (0: "I didn't work hard at all." 10: "I worked as hard as I could."):

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

I find it fair if luck determines income inequality. (0: "I find it completely unfair." 10: "I find it completely fair."):

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

I find it fair if talent determines income inequality. (0: "I find it completely unfair." 10: "I find it completely fair."):

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

I find it fair if how hard people work determines income inequality. (0: "I find it completely unfair." 10: "I find it completely fair."):

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

The harder I work on something, the better I will be at it. (0: "I completely disagree." 10: "I completely agree."):

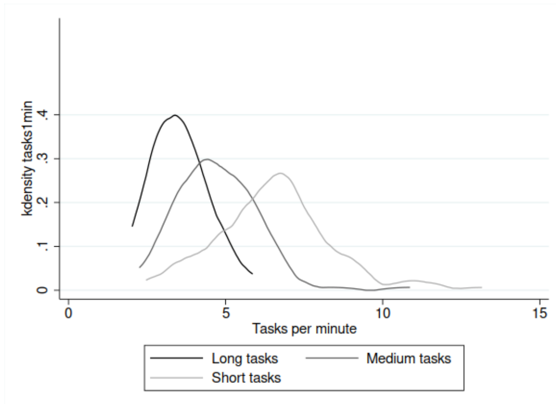
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Talent in an area is something about me that I can't change very much. (0: "I completely disagree." 10: "I completely agree."):

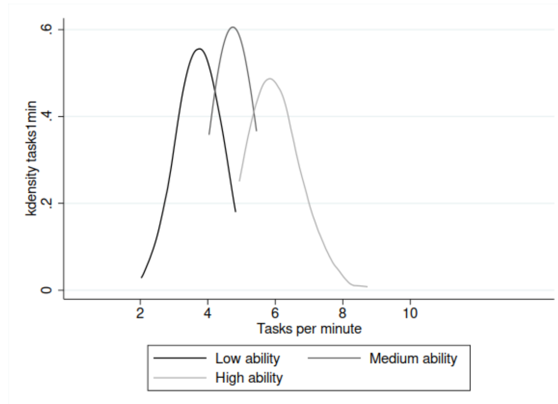
○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Figure A.1: Screenshot of the survey at the end of the experiment

A.2 Distribution of tasks/minute within group



(a) Task length treatment



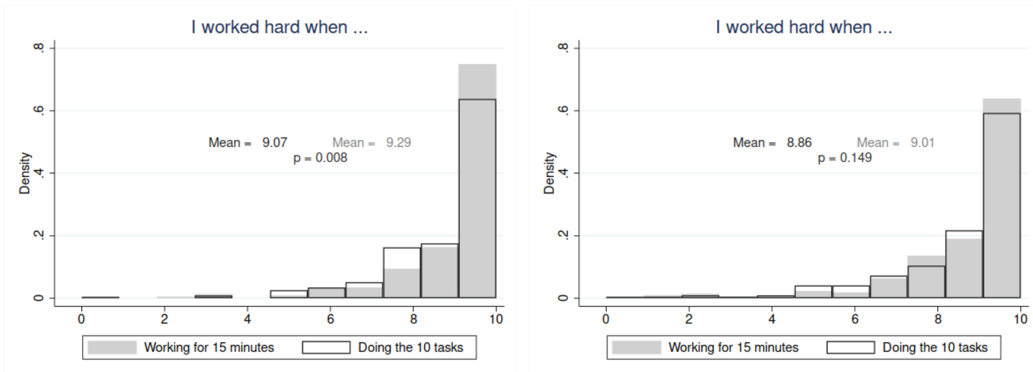
(b) Ability treatment

Figure A.2: Distribution of tasks/minute within group

Note: The figures show the distribution of the individual tasks/minute measures within task length and ability groups. The task length groups are divided by task length – long: 4-letter, medium: 3-letter, short: 2-letter tasks –, and the ability groups are terciles of the tasks/minute measure. There is some overlap in the distributions in the ability treatment because the thresholds between the groups differ across the three sessions.

10.14754/CEU.2022.08

A.3 Production



(a) Task length treatment

(b) Ability treatment

Figure A.3: Self-reported effort level in each stage of the first part

Note: The figures show the distribution of self-reported level of effort in the Production part in the production (15 minutes of work) and when having to do 10 tasks as fast as they can. Participants had to indicate on a 0 to 10 scale how much they agree with the statement "I worked hard when ...".

	Task length experiment Production	Ability experiment Production
Low tasks/min	-20.55***	-13.06***
	(3.369)	(2.028)
High tasks/min	27.39***	14.26***
	(3.283)	(2.016)
Worked hard on production	5.019***	4.285***
	(1.185)	(0.589)
Worked hard on 10 tasks	-1.056	-1.214*
	(1.161)	(0.551)
Constant	68.73***	69.80***
	(2.331)	(1.455)
Observations	257	243

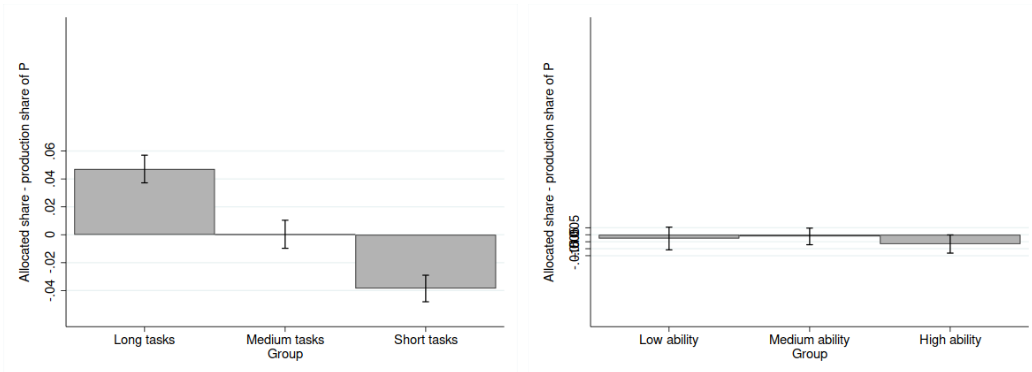
Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.1: Production by group and self-reported effort level

Note: The table shows the effect of own task length / ability group on production, including controls for self-reported effort level in both stages. Self-reported effort level is measured in a 0 to 10 scale. Both effort variables are demeaned. The reference category in the groups in both columns is the medium task length / medium ability group. The constant therefore shows the production of the medium tasks/min groups with average effort level in both stages.

10.14754/CEU.2022.08

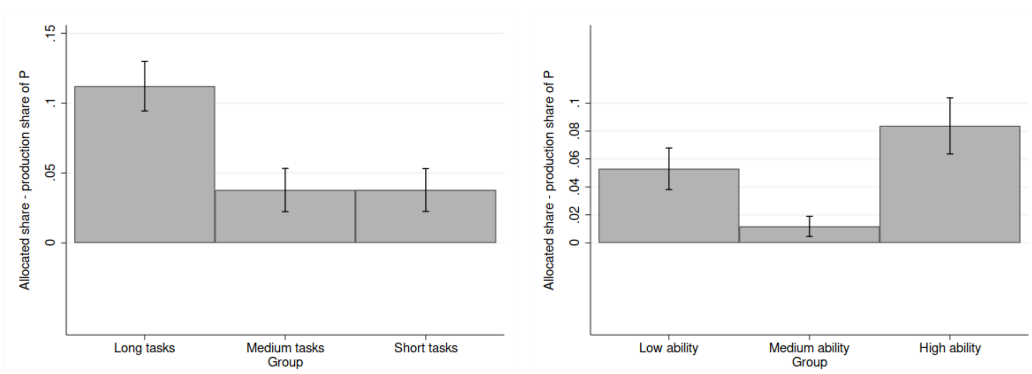
A.4 Redistribution



(a) Task length treatment (b) Ability treatment

Figure A.4: Excess income share allocated to participants by group (spectator decisions)

Note: The figures show the excess income share given to a randomly chosen participant in the pair on top of her production share by spectators, separately by the (task length or ability) group of the participant.



(a) Task length treatment (b) Ability treatment

Figure A.5: Excess income share allocated to self by group (stakeholder decisions)

Note: The figures show the excess income share given to themselves by stakeholders on top of their production share, separately by the (task length or ability) group of the decision-making stakeholder.

10.14754/CEU.2022.08

A.5 Attrition

	Did not come back	Did not come back
Low tasks/min	-0.0410	0.0188
	(0.0535)	(0.0627)
High tasks/min	-0.0488	-0.0492
	(0.0535)	(0.0603)
Task length	-0.0468	-0.0375
experiment	(0.0519)	(0.0590)
Low tasks/min	-0.0200	-0.0707
× Task length experiment	(0.0744)	(0.0862)
High tasks/min	0.0953	0.0773
× Task length experiment	(0.0736)	(0.0847)
Production	-0.00121	-0.000728
	(0.000756)	(0.000968)
Age		-0.000461
		(0.00165)
Female		0.00901
		(0.0380)
Currently studying		-0.00319
		(0.0445)
Employment status		
Full-Time		-0.140
		(0.122)
Not in paid work		-0.201
(e.g. homemaker', 'retired or disabled)		(0.130)
Other		-0.191
		(0.132)
Part-Time		-0.138
		(0.124)
Unemployed (and job		-0.147
seeking)		(0.126)
Constant	0.198^{***}	0.328^{***}
	(0.0373)	(0.127)
Observations	594	451

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Attrition from first to second part

Note: The table shows the effect of features of the experiment and own performance on attrition from the Production to the Redistribution part. The baseline category in the groups is the medium tasks/minute group in the Ability treatment. Age and production is demeaned. In the second column, demographic variables are also included as controls. The number of observations is lower in the second column because demographic data was not available for all participants.

A.6 Robustness checks of reduced-form results

A.6.1 Sample with all demographic controls

	Excess income share to random participant in pair							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equally long tasks								
P had longer tasks	0.0567***	0.0411***	0.0409***	0.0373***				
	(0.00882)	(0.00813)	(0.00819)	(0.00868)				
P had shorter tasks	-0.0549***	-0.0381***	-0.0372***	-0.0416***				
	(0.00810)	(0.00743)	(0.00739)	(0.00860)				
Production share		-0.140***	-0.141***	-0.135***		-0.128***	-0.130***	-0.125***
		(0.0333)	(0.0340)	(0.0353)		(0.0340)	(0.0350)	(0.0368)
Relative difficulty					0.0267***	0.0197***	0.0193***	0.0193***
					(0.00339)	(0.00313)	(0.00315)	(0.00335)
Constant	0.00415	0.0748***	0.0526**	0.0701***	0.00702	0.0713***	0.0466*	0.0639***
	(0.00629)	(0.0162)	(0.0262)	(0.0185)	(0.00600)	(0.0165)	(0.0248)	(0.0184)
Observations	980	980	980	980	980	980	980	980
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Spectator decisions in the task length treatment

Note: The outcome variable is the excess income share given to a randomly chosen participant (Participant 1) in a pair on top of her production share by spectators in the task length treatment. Column 1 controls only for the situation of the participant receiving the income – whether she had longer tasks, equally long tasks, or shorter tasks than the other member of the pair (same as Figure A.4a). Columns 2-4 also control for the production share of Participant 1. Column 3 includes demographic controls: age, gender, whether the spectator was born in the US, whether she has US nationality, student status, and employment status. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the task length groups of P2 and P1 can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

	Excess income share to random participant in pair							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equal ability								
P had lower ability	0.00318	-0.00418	-0.00389	-0.00249				
	(0.00759)	(0.00670)	(0.00663)	(0.00720)				
P had higher ability	0.000524	0.00787	0.00848	0.00987				
	(0.00659)	(0.00569)	(0.00587)	(0.00689)				
Production share		-0.112**	-0.112**	-0.0985*		-0.115**	-0.114**	-0.101*
		(0.0506)	(0.0504)	(0.0541)		(0.0505)	(0.0502)	(0.0542)
Relative difficulty					0.000909	-0.00385	-0.00381	-0.00391
					(0.00329)	(0.00258)	(0.00257)	(0.00281)
Constant	-0.0104*	0.0469*	0.0228	0.0446	-0.00913*	0.0496**	0.0254	0.0484*
	(0.00580)	(0.0248)	(0.0295)	(0.0273)	(0.00462)	(0.0247)	(0.0293)	(0.0270)
Observations	960	960	960	960	960	960	960	960
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.

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

Table A.4: Spectator decisions in the ability treatment

Note: The outcome variable is the excess income share given to a randomly chosen participant (Participant 1) in a pair on top of her production share by spectators in the abiltiy treatment. Column 1 controls only for the situation of the participant receiving the income – whether she had lower ability, equal abiltiy, or higher ability than the other member of the pair (same as Figure A.4b). Columns 2-4 also control for the production share of Participant 1. Column 3 includes demographic controls: age, gender, whether the spectator was born in the US, whether she has US nationality, student status, and employment status. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the ability groups of P2 and P1 can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

	Excess income share to self							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equally long tasks								
P had longer tasks	0.0838***	0.0551**	0.0531***	0.0224**				
	(0.0231)	(0.0221)	(0.0196)	(0.00951)				
P had shorter tasks	-0.0236	0.00406	0.00551	0.00702				
	(0.0145)	(0.0164)	(0.0147)	(0.0113)				
Production share		-0.272***	-0.274***	-0.326***		-0.279***	-0.275***	-0.315***
		(0.0683)	(0.0687)	(0.0602)		(0.0713)	(0.0718)	(0.0638)
Relative difficulty					0.0242***	0.0102	0.0107	0.00510*
					(0.00675)	(0.00718)	(0.00653)	(0.00305)
Constant	0.0319	0.165***	0.252***	0.221***	0.0567**	0.192***	0.273***	0.225***
	(0.0260)	(0.0400)	(0.0902)	(0.0296)	(0.0273)	(0.0425)	(0.0930)	(0.0322)
Observations	960	960	960	960	960	960	960	960
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Stakeholder decisions in the task length treatment

Note: The outcome variable is the excess income share given to self in a pair on top of her production share by stakeholders in the task length treatment. Column 1 controls only for the situation of the decision-maker stakeholder – whether she had longer tasks, equally long tasks, or shorter tasks than the other member of the pair (same as Figure A.5a). Columns 2-4 also control for the production share of the stakeholder. Column 3 includes demographic controls: age, gender, whether she was born in the US, whether she has US nationality, student status, and employment status. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the task length groups of the other pair member and the stakeholder can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

10.14754/CEU.2022.08

	Excess income share to self							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Situation, ref. equal ability								
P had lower ability	-0.00130	-0.0227	-0.0148	0.00659				
	(0.0188)	(0.0253)	(0.0190)	(0.00834)				
P had higher ability	0.00436	0.0315*	0.0284*	0.0130				
	(0.0192)	(0.0164)	(0.0147)	(0.0115)				
Production share		-0.399**	-0.383**	-0.168***		-0.421*	-0.393**	-0.168***
		(0.196)	(0.160)	(0.0552)		(0.219)	(0.171)	(0.0541)
Relative difficulty					-0.000954	-0.0182	-0.0141	-0.00240
					(0.00886)	(0.0135)	(0.00968)	(0.00402)
Constant	0.0444	0.247**	0.288**	0.119***	0.0453	0.262**	0.298***	0.126***
	(0.0304)	(0.110)	(0.111)	(0.0297)	(0.0301)	(0.114)	(0.107)	(0.0274)
Observations	880	880	880	880	880	880	880	880
Participant fixed effect	no	no	no	yes	no	no	no	yes
Demographic controls	no	no	yes	no	no	no	yes	no
Session fixed effect	yes	yes	yes	no	yes	yes	yes	no

Standard errors are clustered on participant level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Stakeholder decisions in the task length treatment

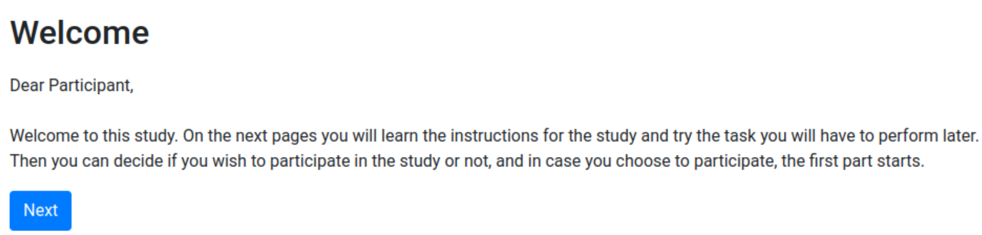
Note: The outcome variable is the excess income share given to self in a pair on top of her production share by stakeholders in the ability treatment. Column 1 controls only for the situation of the decision-maker stakeholder – whether she had lower, equal, or higher ability than the other member of the pair (same as Figure A.5b). Columns 2-4 also control for the production share of the stakeholder. Column 3 includes demographic controls: age, gender, whether she was born in the US, whether she has US nationality, student status, and employment status. Column 4 adds participant fixed effects. In columns 5-8, relative difficulty is a continuous measure that is the difference between the number of tasks the ability groups of the other pair member and the stakeholder can do within a minute. Column 6 adds production share, column 7 demographic controls, and column 8 participant fixed effects.

10.14754/CEU.2022.08

A.7 Experiment instructions

A.7.1 Production

The example task in the screenshots is a 4-letter task in the task length treatment, but the Production part instructions were the same for all participants across treatments.



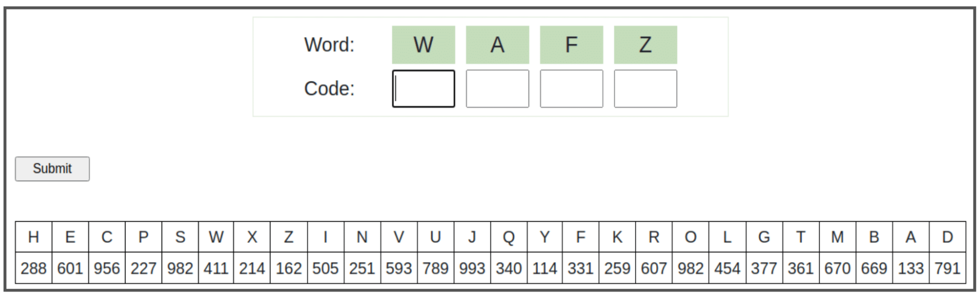
10.14754/CEU.2022.08

Instructions

This study will take place entirely online, in two parts. If you decide to take part in the study, the first part will start immediately after you give consent to participate. You can complete the second part anytime tomorrow between 6 AM and midnight (Pacific Time). You will receive a **completion bonus of £2.50** for completing the first part, **plus £1.00** if you complete the second part as well. You can earn **additional income** that will depend on your performance in the first part and your decisions or on other participants' decisions in the second part. The average additional income will be **around £3.00.** The first part will take around 25-30 minutes, and the second part around 5-10 minutes to complete.

First part - TODAY

In the first part you will have to do a simple task. In this task you have to encrypt letter combinations (= "words") into numbers, based on the key you see at the bottom of the page. For reference, see **a screenshot of the task** below.



Word:	W	A	F	Z
Code:				

Submit

H	E	C	P	S	W	X	Z	I	N	V	U	J	Q	Y	F	K	R	O	L	G	T	M	B	A	D
288	601	956	227	982	411	214	162	505	251	593	789	993	340	114	331	259	607	982	454	377	361	670	669	133	791

In this example the letter "W" corresponds to the number 411, "A" to 133, "F" to 331, and "Z" to 162, so you have to type 411, 133, 331, and 162 into the boxes under the letters, respectively. You can submit your response by clicking the Submit button below the task. If any of the entries are incorrect, the computer will tell you the number of entries that are wrong but not which ones are wrong. Then you will be able to revise your solution. Once the correct numbers are entered, after clicking the Submit button, you will see a new word to encrypt. After each correctly solved task the computer generates a new encryption key, too. The order of the letters in the key is also shuffled between the tasks.

When a new word is generated, **the first box will become active,** so you can start entering the first number. The fastest way to navigate from one box to the next is to use the **tabulator key (Tab)** on your keyboard. From the last box you can navigate to the Submit button directly with the Tab key and hit Enter. The tab key looks like this on your keyboard on PC (1) and on Mac (2 or 3):

(1) (2) (3)

On the next page you can try the task you will do in the experiment. You will be automatically navigated to the following page after completing the task or after 2 minutes.

Next

10.14754/CEU.2022.08

Try task

Time left to complete this page: **1:50**

Word:	G	A	R	B
Code:				

Submit

D	F	U	P	C	X	O	L	S	Y	M	B	A	H	G	W	R	V	J	I	N	E	Q	Z	K	T
124	227	474	628	521	484	448	672	767	853	598	831	440	631	993	695	562	922	733	303	879	526	735	788	794	220

Reminder: The fastest way to navigate from one box to the next is to use the **tabulator key (Tab)** on your keyboard. From the last box you can navigate to the Submit button directly with the Tab key and hit Enter. The tab key looks like this on your keyboard on PC (1) and on Mac (2 or 3):

(1) (2) (3)

Instructions - continued

First part - TODAY

The first part will take around 25-30 minutes to complete and will have **two main stages**. In both stages you will have to perform the same task you just tried. You will receive £2.50 for completing this part.

1. 10 tasks

First, I will ask you to do **10 tasks, as fast as you can,** to measure how quickly you can perform the tasks. Although you do not receive payment for this part in particular, it will serve as **information about how fast these tasks can be done.** It may also affect your additional income (on top of the completion bonuses) determined in the second part. You will have 5 minutes for this part. You will only be able to continue with the study if you **complete the 10 tasks within the 5 minutes.** Before doing the 10 tasks you will be able to practice for 2 minutes.

2. Production

In this part, you will have to **do the task for 15 minutes.** The **number of tasks you correctly completed** within these 15 minutes will be **your production.**

Earnings from the first part

You will earn **10 tokens for each completed task** in the Production stage.

Note that this is **not how much you will eventually earn in addition to the completion bonuses**. Your final additional income may change according to your decisions or other participants' **decisions in the second part**.

Second part - TOMORROW

The second part will take place tomorrow, and you can complete it anytime between 6 AM and midnight (Pacific Time). It will take around 5-10 minutes to complete. You will receive a completion bonus of £1.00 for completing this part. You will receive an **invitation tomorrow via the Prolific emailing system with the link to the second part**. If you complete the first part, when entering the second part the computer will automatically recognize you by your Prolific ID and let you start.

At the beginning of this part, you will be **randomly paired with another participant** who also completed the first part. In this part you will have to make **decisions about the distribution of earnings from the first part** within pairs. Your final additional income, that is on top of the fixed completion bonuses, will therefore depend on your decisions or on other participants' decisions made in the second part. The **completion bonuses** for the first and the second parts **will not be affected in any way by either your or other participants' decisions. The decisions will affect only the additional income you will earn.** The final additional earnings will be converted to British Pounds at the end of the study with **250 tokens = £1.00**. You will receive more detailed instructions about the second part at the beginning of the second part.

Payment for the study

You will receive **all payments within 3 days after the second part** via the Prolific payment system. **If you only complete the first part, you will receive £2.50 within 3 days after the corresponding second part.** Due to the interactive nature of the study we will have to wait for all participants' answers to draw the final payments. Because of this, **you will only learn how much you earned in total when you receive your payments for the study.**

On the next page you can test your comprehension of the study and you can decide if you wish to participate. You will see the instructions again under the consent form.

Further Information

This study is conducted by Luca Flora Drucker and financed by Central European University. If you accept to participate in the study, you may still change your mind and quit at any time. However, please note that you only receive the completion bonus for the first part and the completion bonus for the second part and the additional income if you complete the first and the second parts, respectively.

Participation in this study is not associated with any foreseeable risk or benefit. Your answers will be collected confidentially and anonymously (the researcher will not be able to link decisions and participants' identity beyond the Prolific ID provided). At the data analysis stage your Prolific ID will be changed to a random identifying number, and the Prolific IDs will be deleted. In case the results of the study are published, there can be no references to your identity. Data anonymity is guaranteed.

This study received a research ethics approval from the Ethical Research Committee of Central European University.

If you have any questions or concerns regarding this study, please contact me at lucafloradrucker.research@gmail.com.

Next

10.14754/CEU.2022.08

Comprehension check and consent form

Please answer the following questions before you decide if you wish to participate, to make sure you understand the details of the study. You can see the instructions again at the bottom of this page for reference. If you give an incorrect answer to any of the questions and click the Next button, the computer will notify you that the answer is wrong.

When is the second part of the study?

- Today.
- Tomorrow.
- In two days.

If you only complete the first part, what will be your payment?

- The completion bonus for the first part plus the payment after the number of tasks you do.
- Only the completion bonus for the first part.
- Nothing.

How will you earn additional income on top of the completion bonuses?

- The additional income will be based on the tasks you perform in the production stage, but will be determined by decisions in the second part.
- The number of tasks you perform in the production stage will be your additional income.
- The additional income will entirely be determined by luck.

Consent Form

Do you wish to participate in the study?

- Yes
- No

Next

Practice

Welcome to the study! Thank you for taking part.
Now you can **practice the task for 2 minutes.**

Next

10 tasks

Now please do **10 tasks as fast as you can,** to measure how fast you can solve the tasks. You will have 5 minutes for this part. You have to finish the 10 tasks within the 5 minutes **to be able to continue the study.**

Next

Beginning of production

Thank you for completing the 10 tasks. Next, I will ask you to do the same task for **15 minutes.** The timer starts when you click the Next button on this page. The **number of tasks you do within the 15 minutes will be your production.** You will earn 10 tokens per task correctly done. The resulting amount will be your **income to distribute in the second part.** You will only receive information about your exact production at the beginning of the second part.

Please click on the Next button to proceed.

Next

End of first part

This is the end of the first part. Thank you for participating!

Tomorrow you will receive an invitation to the second part, which will be open between 6 AM and midnight (Pacific Time).

Please click HERE to **go back to Prolific, proving that you completed this part of the study.** Alternatively, you can go back to Prolific, and manually enter the following **completion code:**

10.14754/CEU.2022.08

A.7.2 Redistribution – task length spectators

Instructions for Part 2

Your performance and more information about the first part

In the first part you performed 45 tasks, therefore your earnings from the first part are 450 tokens.

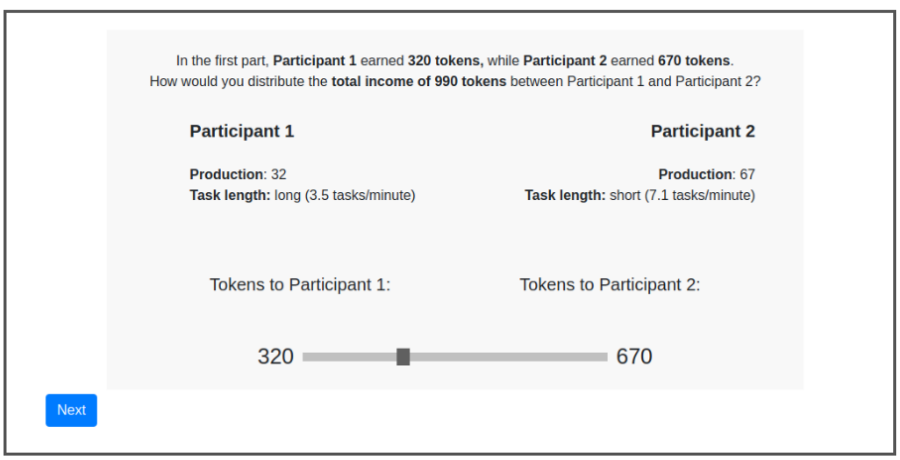
During the first part, you had to work on **3-letter tasks**. However, only about a third of the other participants had 3-letter tasks as well, another **third of participants had to work on 2-letter and the other third on 4-letter tasks**. The task length was **randomly selected** at the beginning of the study for each participant.

At the beginning of the first part, you had to do 10 tasks, as fast as you could. Based on how fast all participants solved the 10 tasks, the computer calculated **how many tasks participants were able to solve within a minute** on average in each task length group. These are **6.2 tasks per minute for 2-letter tasks, 4.5 for 3-letter tasks, and 3.2 for 4-letter tasks**.

Instructions for the second part

For this part, the computer **paired you randomly with another participant** who also completed the first part yesterday. Besides assigning you to another participant, the computer assigned **another pair of participants** from the first part to your group of two. On the following pages, you will have to **decide how to distribute the joint earnings of these two other participants**. Your partner will also make distribution decisions over the joint earnings of these two other participants. Consequently, another pair of participants will distribute your and your partner's earnings.

For making the decisions you will learn both participants' **earnings** and **production** from the first part, and both participants' **task length**. You will also be reminded about **the average number of tasks per minute** participants in the task length group were able to do **when doing the 10 tasks,** so how easy or difficult it was for both participants to solve the tasks. This is an example of the decision screen (the numbers in the screenshot are hypothetical):



When you arrive to the decision page, **you have to click on the slider to reveal the handle.** If you move the handle, you see on the left and the right of the slider the amounts to the first and the second participant that correspond to the current position of the handle. You will be able to move the handle in 5-token increments. Note that if you want to e.g., **increase the payment to the participant on the left**, you have to move the **handle to the right.** This way the number of tokens given to the participant on the left increases, and the number of tokens to the participant on the right decreases. To make a decision, move the handle to the position of the chosen allocation, and click on the Next button.

You will make **10 decisions** in total, of which only **one corresponds to distribution between the actual two participants assigned to you,** and the others correspond to potential other pairings of participants from the first session that did not take place. We cannot tell you which one is the actual decision, so please consider each decision as it were to be implemented at the end of the study.

10.14754/CEU.2022.08

After the decisions, I will ask you to fill in a short survey and then the study will end.

After every participant finishes the second part, the computer will select **one decision for each pair randomly from the two decisions made for that pair.** That allocation decision will be implemented, so your final payment will be the income given to you in the allocation decision **made by one of the members of another pair.** Income will be converted to British Pounds with 250 tokens = £1.00. You will learn your final payment when you receive all payments through the Prolific payment system, within three days.

By clicking the Next button you will proceed to the decisions.

Next

10.14754/CEU.2022.08

A.7.3 Redistribution – task length stakeholders

Instructions for Part 2

Your performance and more information about the first part

In the first part you performed 45 tasks, therefore your earnings from the first part are 450 tokens.

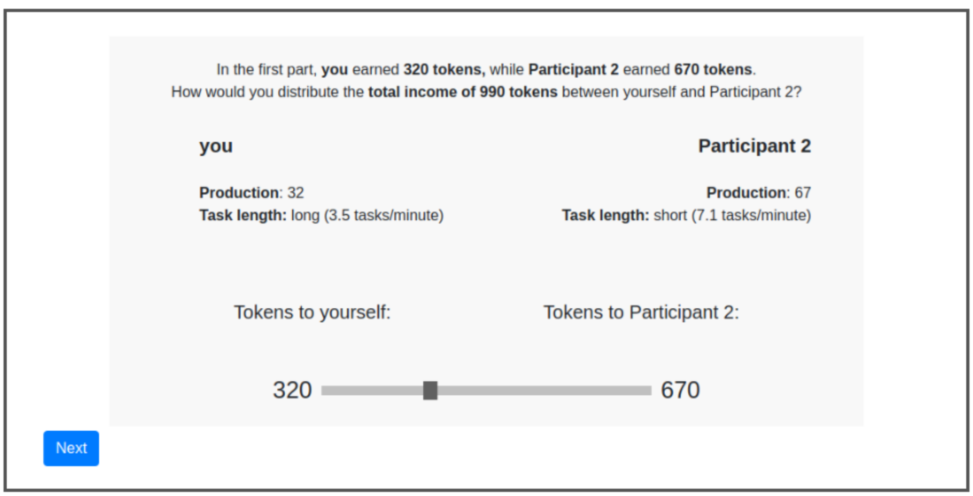
During the first part, you had to work on **4-letter tasks**. However, only about a third of the other participants had 4-letter tasks as well, another **third of participants had to work on 2-letter and the other third on 3-letter tasks**. The task length was **randomly selected** at the beginning of the study for each participant.

At the beginning of the first part, you had to do 10 tasks, as fast as you could. Based on how fast all participants solved the 10 tasks, the computer calculated **how many tasks participants were able to solve within a minute** on average in each task length group. These are **6.2 tasks per minute for 2-letter tasks, 4.5 for 3-letter tasks, and 3.2 for 4-letter tasks**.

Instructions for the second part

For this part, the computer **paired you randomly with another participant** who also completed the first part yesterday. On the following pages, you will have to **decide how to distribute the joint earnings of you and your partner**. Your partner will also make distribution decisions over your joint earnings.

For making the decisions you will learn your partner's **earnings** and **production** from the first part, and your and your partner's **task length**. You will also be reminded about **the average number of tasks per minute** participants in the task length group were able to do **when doing the 10 tasks,** so how easy or difficult it was for you and your partner to solve the tasks. This is an example of the decision screen (the numbers in the screenshot are hypothetical):



When you arrive to the decision page, **you have to click on the slider to reveal the handle.** If you move the handle, you see on the left and the right of the slider the amounts to you and to your partner that correspond to the current position of the handle. You will be able to move the handle in 5-token increments. Note that if you want to e.g., **increase the payment to the participant on the left**, you have to move the **handle to the right.** This way the number of tokens given to the participant on the left increases, and the number of tokens to the participant on the right decreases. To make a decision, move the handle to the position of the chosen allocation, and click on the Next button.

You will make **10 decisions** in total, of which only **one corresponds to distribution between you and your actual partner,** and the others correspond to potential other pairings of participants from the first session that did not take place. We cannot tell you which one is the actual decision, so please consider each decision as it were to be implemented at the end of the study. Note that if in the particular decision you produced less than your partner, your data will be on the **left-hand side**, while if you produced more than your partner, your data will be on the **right-hand side** on the screen.

10.14754/CEU.2022.08

After the decisions, I will ask you to fill in a short survey and then the study will end.

After every participant finishes the second part, the computer will select **one decision for each pair randomly from the two decisions made for that pair.** That allocation decision will be implemented, so your final payment will be the income given to you in the allocation decision **made by you or your partner.** Income will be converted to British Pounds with 250 tokens = £1.00. You will learn your final payment when you receive all payments through the Prolific payment system, within three days.

By clicking the Next button you will proceed to the decisions.

Next

A.7.4 Redistribution – ability spectators

Instructions for Part 2

Your performance and more information about the first part

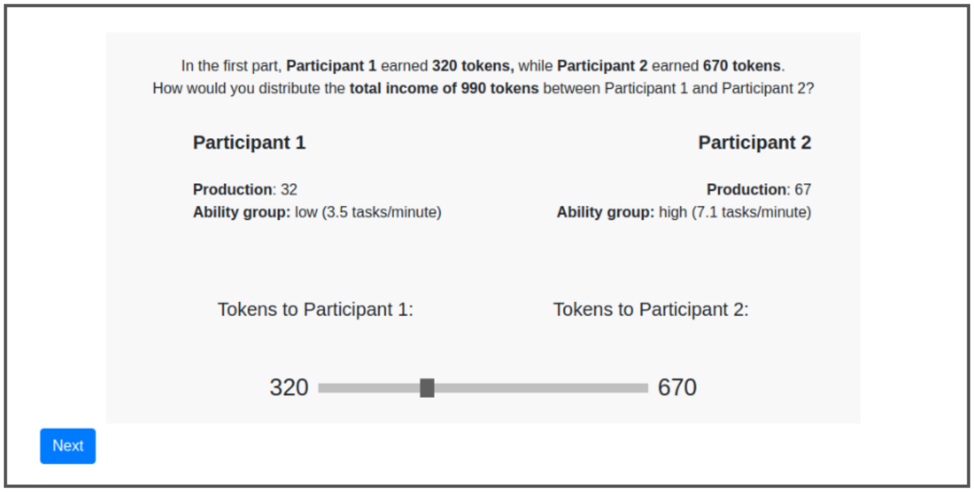
In the first part you performed 45 tasks, therefore your earnings from the first part are 450 tokens.

At the beginning of the first part, you had to do 10 tasks, as fast as you could. Based on how fast all participants solved the 10 tasks, the computer created **three ability groups** that measure **how good participants are in this particular task**. From the time it took for the 10 tasks it also calculated **how many tasks participants were able to solve within a minute** on average in each ability group. These are **3.2 for the low ability group, 4.5 for the medium, and 6.2 for the high ability group.** Based on how fast you solved the 10 tasks, you fell into the third, high ability group in this task.

Instructions for the second part

For this part, the computer **paired you randomly with another participant** who also completed the first part yesterday. Besides assigning you to another participant, the computer assigned **another pair of participants** from the first part to your group of two. On the following pages, you will have to **decide how to distribute the joint earnings of these two other participants**. Your partner will also make distribution decisions over the joint earnings of these two other participants. Consequently, another pair of participants will distribute your and your partner's earnings.

For making the decisions you will learn both participants' **earnings** and **production** from the first part, and both participants' **ability group**. You will also be reminded about **the average number of tasks per minute** participants in the ability group were able to do **when doing the 10 tasks,** so how easy or difficult it was for both participants to solve the tasks. This is an example of the decision screen (the numbers in the screenshot are hypothetical):



When you arrive to the decision page, **you have to click on the slider to reveal the handle.** If you move the handle, you see on the left and the right of the slider the amounts to the first and the second participant that correspond to the current position of the handle. You will be able to move the handle in 5-token increments. Note that if you want to e.g., **increase the payment to the participant on the left**, you have to move the **handle to the right.** This way the number of tokens given to the participant on the left increases, and the number of tokens to the participant on the right decreases. To make a decision, move the handle to the position of the chosen allocation, and click on the Next button.

You will make **10 decisions** in total, of which only **one corresponds to distribution between the actual two participants assigned to you,** and the others correspond to potential other pairings of participants from the first session that did not take place. We cannot tell you which one is the actual decision, so please consider each decision as it were to be implemented at the end of the study.

After the decisions, I will ask you to fill in a short survey and then the study will end.

After every participant finishes the second part, the computer will select **one decision for each pair randomly from the two decisions made for that pair.** That allocation decision will be implemented, so your final payment will be the income given to you in the allocation decision **made by one of the members of another pair.** Income will be converted to British Pounds with 250 tokens = £1.00. You will learn your final payment when you receive all payments through the Prolific payment system, within three days.

By clicking the Next button you will proceed to the decisions.

Next

10.14754/CEU.2022.08

A.7.5 Redistribution – ability stakeholders

Instructions for Part 2

Your performance and more information about the first part

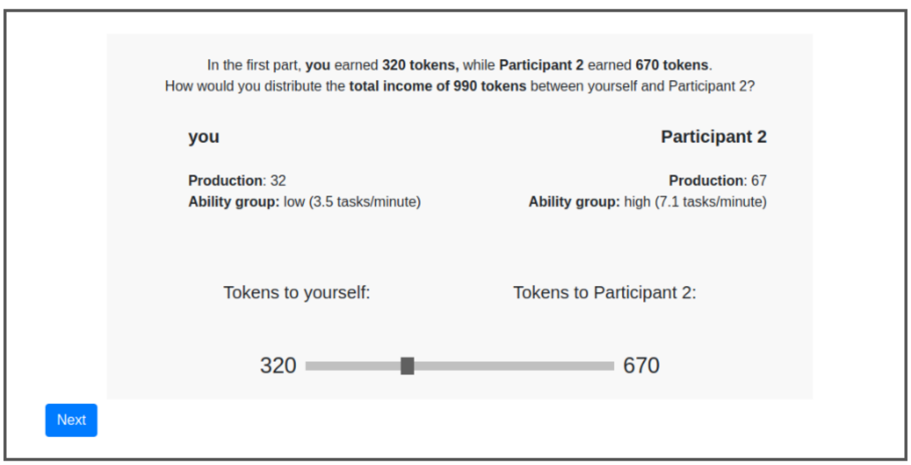
In the first part you performed 45 tasks, therefore your earnings from the first part are 450 tokens.

At the beginning of the first part, you had to do 10 tasks, as fast as you could. Based on how fast all participants solved the 10 tasks, the computer created **three ability groups** that measure **how good participants are in this particular task**. From the time it took for the 10 tasks it also calculated **how many tasks participants were able to solve within a minute** on average in each ability group. These are **3.2 for the low ability group, 4.5 for the medium, and 6.2 for the high ability group.** Based on how fast you solved the 10 tasks, you fell into the third, high ability group in this task.

Instructions for the second part

For this part, the computer **paired you randomly with another participant** who also completed the first part yesterday. On the following pages, you will have to **decide how to distribute the joint earnings of you and your partner**. Your partner will also make distribution decisions over your joint earnings.

For making the decisions you will learn your partner's **earnings** and **production** from the first part, and your and your partner's **ability group**. You will also be reminded about **the average number of tasks per minute** participants in the ability group were able to do **when doing the 10 tasks,** so how easy or difficult it was for you and your partner to solve the tasks. This is an example of the decision screen (the numbers in the screenshot are hypothetical):



When you arrive to the decision page, **you have to click on the slider to reveal the handle.** If you move the handle, you see on the left and the right of the slider the amounts to you and to your partner that correspond to the current position of the handle. You will be able to move the handle in 5-token increments. Note that if you want to e.g., **increase the payment to the participant on the left**, you have to move the **handle to the right.** This way the number of tokens given to the participant on the left increases, and the number of tokens to the participant on the right decreases. To make a decision, move the handle to the position of the chosen allocation, and click on the Next button.

You will make **10 decisions** in total, of which only **one corresponds to distribution between you and your actual partner,** and the others correspond to potential other pairings of participants from the first session that did not take place. We cannot tell you which one is the actual decision, so please consider each decision as it were to be implemented at the end of the study. Note that if in the particular decision you produced less than your partner, your data will be on the **left-hand side**, while if you produced more than your partner, your data will be on the **right-hand side** on the screen.

10.14754/CEU.2022.08

After the decisions, I will ask you to fill in a short survey and then the study will end.

After every participant finishes the second part, the computer will select **one decision for each pair randomly from the two decisions made for that pair.** That allocation decision will be implemented, so your final payment will be the income given to you in the allocation decision **made by you or your partner.** Income will be converted to British Pounds with 250 tokens = £1.00. You will learn your final payment when you receive all payments through the Prolific payment system, within three days.

By clicking the Next button you will proceed to the decisions.

Next

10.14754/CEU.2022.08

B Appendix for Chapter 2

B.1 Choices in batches 1 and 2 of the first experiment

Batch 1:

1. 20 large matrices in Session 2 or 5 large matrices in Session 2 + \$0
2. 23 large matrices in Session 1 or 5 large matrices in Session 2 + \$0
3. 20 large matrices in Session 2 or 19 large matrices in Session 2 with $p = 0.9$ + \$0.20
4. 23 large matrices in Session 1 or 19 large matrices in Session 2 with $p = 0.9$ + \$0.20

Batch 2:

1. 22 large matrices in Session 2 or 15 large matrices in Session 2 + \$1.80
2. 25 large matrices in Session 1 or 15 large matrices in Session 2 + \$1.80
3. 22 large matrices in Session 2 or 15 large matrices in Session 2 with $p = 0.9$ + \$0.60
4. 25 large matrices in Session 1 or 15 large matrices in Session 2 with $p = 0.9$ + \$0.60

B.2 Debrief survey statistics

Table B.1 shows summary statistics from the debrief survey for those who gave consent but not completed the experiment (attritors) and those who completed the whole experiment (completors), separately. Most participants found the instructions clear, and found the large matrix significantly less pleasant than the small matrix. Interestingly, completors hate the large matrix more than attritors, but the two groups are similar in other aspects. They work on average $\sim$ 20 hours weekly on MTurk, and earn $\sim$ 132 dollars per week from this work.

10.14754/CEU.2022.08

Table B.1: Summary statistics from debrief survey

	(1) Attritors		(2) Completors		(3) Difference	
	mean	sd	mean	sd	diff	t
Instructions are clear	0.96	(0.20)	0.96	(0.20)	0.00	(0.00)
How pleasant is small matrix (0-10)	5.16	(2.83)	4.85	(2.15)	0.31	(0.71)
How pleasant is large matrix (0-10)	4.78	(3.14)	3.43	(2.33)	1.35***	(2.76)
How much worse is large matrix	0.39	(1.90)	1.42	(1.86)	-1.03***	(-3.31)
Patience (0-10)	7.37	(2.32)	7.22	(1.97)	0.14	(0.39)
Weekly working hours on MTurk	18.82	(15.19)	20.29	(13.93)	-1.47	(-0.60)
Weekly earnings on MTurk	118.43	(89.78)	135.63	(99.60)	-17.20	(-1.13)
Observations	49		147		196	

Note: Attritors are those who gave consent to participate, but did not finish the experiment, Completors are who completed all aspects of the experiment. In rows 2-3 (”How pleasant is small matrix” and ”How pleasant is large matrix”) 0 means very unpleasant and 10 means very pleasant compared to other tasks they do on MTurk. Row 4 is calculated as row 2 - row 3 individually. In row 5 (”Patience”) 0 means very impatient and 10 means very patient. Weekly earnings on MTurk are in US dollars.
$^{*}p < 0.1$ $^{**}p < 0.05$ $^{***}p < 0.01$

10.14754/CEU.2022.08

B.3 Choices in the second experiment

B.3.1 Risk batches

The tasks are either all Greek tasks or all matrices within a batch. All tasks are additional to the 10 required tasks in the session. In the decisions about work two weeks ahead, the dates are ”Monday in two weeks” and ”Thursday in two weeks”.

Batch 1

1. 20 tasks on Thursday or 13 tasks on Thursday + €2.20
2. 20 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €2.20
3. 19 tasks today or 13 tasks on Thursday + €1.20
4. 19 tasks today or 10 tasks on Thursday with $p = 0.9$ + €1.20

Batch 2

1. 18 tasks on Thursday or 14 tasks on Thursday + €1.80
2. 18 tasks on Thursday or 12 tasks on Thursday with $p = 0.9$ + €1.80
3. 21 tasks today or 14 tasks on Thursday + €0.80
4. 21 tasks today or 12 tasks on Thursday with $p = 0.9$ + €0.80

Batch 3

1. 22 tasks on Thursday or 16 tasks on Thursday + €1.00
2. 22 tasks on Thursday or 13 tasks on Thursday with $p = 0.9$ + €1.00
3. 25 tasks today or 16 tasks on Thursday + €0.80
4. 25 tasks today or 13 tasks on Thursday with $p = 0.9$ + €0.80

Batch 4

1. 25 tasks on Thursday or 12 tasks on Thursday + €1.60
2. 25 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €1.60
3. 23 tasks today or 12 tasks on Thursday + €1.00
4. 23 tasks today or 10 tasks on Thursday with $p = 0.9$ + €1.00

Batch 5

1. 11 tasks on Thursday or 11 tasks on Thursday + €1.20
2. 11 tasks on Thursday or 11 tasks on Thursday with $p = 0.9$ + €1.20

3. 11 tasks today or 11 tasks on Thursday + €1.20
4. 11 tasks today or 11 tasks on Thursday with $p = 0.9$ + €1.20

Batch 6

1. 17 tasks on Thursday or 10 tasks on Thursday + €0.40
2. 17 tasks on Thursday or 10 tasks on Thursday with $p = 0.9$ + €0.40
3. 10 tasks today or 10 tasks on Thursday + €0.20
4. 10 tasks today or 10 tasks on Thursday with $p = 0.9$ + €0.20

Batch 7

1. 15 tasks on Thursday or 15 tasks on Thursday + €1.00
2. 15 tasks on Thursday or 15 tasks on Thursday with $p = 0.9$ + €1.00
3. 15 tasks today or 15 tasks on Thursday + €1.00
4. 15 tasks today or 15 tasks on Thursday with $p = 0.9$ + €1.00

Batch 8

1. 12 tasks on Thursday or 8 tasks on Thursday + €0.20
2. 12 tasks on Thursday or 8 tasks on Thursday with $p = 0.9$ + €0.20
3. 12 tasks today or 8 tasks on Thursday + €0.00
4. 12 tasks today or 8 tasks on Thursday with $p = 0.9$ + €0.00

B.3.2 Matrix-versus-Greek batches

All tasks are additional to the 10 required matrix or Greek tasks in the session. In the decisions about work two weeks ahead, the dates are ”Monday in two weeks” and ”Thursday in two weeks”.

Batch 1

1. 20 matrix (Greek) tasks on Thursday or 12 matrix (Greek) tasks on Thursday + €1.80
2. 20 matrix (Greek) tasks on Thursday or 9 Greek (matrix) tasks on Thursday + €1.80
3. 18 matrix (Greek) tasks today or 12 matrix (Greek) tasks on Thursday + €1.20
4. 18 matrix (Greek) tasks today or 9 Greek (matrix) tasks on Thursday + €1.20

Batch 2

1. 19 matrix (Greek) tasks on Thursday or 13 matrix (Greek) tasks on Thursday + €1.60
2. 19 matrix (Greek) tasks on Thursday or 10 Greek (matrix) tasks on Thursday + €1.60
3. 23 matrix (Greek) tasks today or 13 matrix (Greek) tasks on Thursday + €0.60
4. 23 matrix (Greek) tasks today or 10 Greek (matrix) tasks on Thursday + €0.60

Batch 3

1. 25 matrix (Greek) tasks on Thursday or 14 matrix (Greek) tasks on Thursday + €1.00
2. 25 matrix (Greek) tasks on Thursday or 11 Greek (matrix) tasks on Thursday + €1.00
3. 22 matrix (Greek) tasks today or 14 matrix (Greek) tasks on Thursday + €0.80
4. 22 matrix (Greek) tasks today or 11 Greek (matrix) tasks on Thursday + €0.80

Batch 4

1. 21 matrix (Greek) tasks on Thursday or 15 matrix (Greek) tasks on Thursday + €1.40
2. 21 matrix (Greek) tasks on Thursday or 12 Greek (matrix) tasks on Thursday + €1.40
3. 24 matrix (Greek) tasks today or 15 matrix (Greek) tasks on Thursday + €1.20
4. 24 matrix (Greek) tasks today or 12 Greek (matrix) tasks on Thursday + €1.20

Batch 5

1. 11 matrix (Greek) tasks on Thursday or 11 matrix (Greek) tasks on Thursday + €0.80
2. 11 matrix (Greek) tasks on Thursday or 8 Greek (matrix) tasks on Thursday + €0.80
3. 11 matrix (Greek) tasks today or 11 matrix (Greek) tasks on Thursday + €0.80
4. 11 matrix (Greek) tasks today or 8 Greek (matrix) tasks on Thursday + €0.80

Batch 6

1. 16 matrix (Greek) tasks on Thursday or 16 matrix (Greek) tasks on Thursday + €0.60
2. 16 matrix (Greek) tasks on Thursday or 16 Greek (matrix) tasks on Thursday + €0.60

3. 16 matrix (Greek) tasks today or 16 matrix (Greek) tasks on Thursday + €0.60
4. 16 matrix (Greek) tasks today or 16 Greek (matrix) tasks on Thursday + €0.60

Batch 7

1. 20 matrix (Greek) tasks on Thursday or 14 matrix (Greek) tasks on Thursday + €2.00
2. 20 matrix (Greek) tasks on Thursday or 14 Greek (matrix) tasks on Thursday + €2.00
3. 20 matrix (Greek) tasks today or 14 matrix (Greek) tasks on Thursday + €1.20
4. 20 matrix (Greek) tasks today or 14 Greek (matrix) tasks on Thursday + €1.20

Batch 8

1. 15 matrix (Greek) tasks on Thursday or 15 matrix (Greek) tasks on Thursday + €1.20
2. 15 matrix (Greek) tasks on Thursday or 10 Greek (matrix) tasks on Thursday + €1.20
3. 15 matrix (Greek) tasks today or 15 matrix (Greek) tasks on Thursday + €1.20
4. 15 matrix (Greek) tasks today or 10 Greek (matrix) tasks on Thursday + €1.20

B.3.3 Follow-up risk

All tasks are matrix tasks.

Batch 1

1. 11 hard tasks on Thursday or 11 hard tasks on Thursday + €1.20
2. 11 hard tasks on Thursday or 11 hard tasks on Thursday with $p = 0.9$ + €1.20
3. 11 hard tasks today or 11 hard tasks on Thursday + €1.20
4. 11 hard tasks today or 11 hard tasks on Thursday with $p = 0.9$ + €1.20

Batch 2

1. 17 hard tasks on Thursday or 10 hard tasks on Thursday + €0.40
2. 17 hard tasks on Thursday or 10 hard tasks on Thursday with $p = 0.9$ + €0.40
3. 17 hard tasks today or 10 hard tasks on Thursday + €0.20
4. 17 hard tasks today or 10 hard tasks on Thursday with $p = 0.9$ + €0.20

10.14754/CEU.2022.08

Batch 3

1. 15 hard tasks on Thursday or 15 hard tasks on Thursday + €1.00
2. 15 hard tasks on Thursday or 15 hard tasks on Thursday with $p = 0.9$ + €1.00
3. 15 hard tasks today or 15 hard tasks on Thursday + €1.00
4. 15 hard tasks today or 15 hard tasks on Thursday with $p = 0.9$ + €1.00

Batch 4

1. 12 hard tasks on Thursday or 8 hard tasks on Thursday + €0.20
2. 12 hard tasks on Thursday or 8 hard tasks on Thursday with $p = 0.9$ + €0.20
3. 12 hard tasks today or 8 hard tasks on Thursday + €0.20
4. 12 hard tasks today or 8 hard tasks on Thursday with $p = 0.9$ + €0.20

B.3.4 Ruling out reference dependence

Let us regard one general batch:

1. x_1 in three days + X or x_2 in three days + \$0
2. x_3 today + X' or x_2 in three days + \$0
3. x_1 in in three days + Y or x_4 in three days with $p = 0.9$ + \$$z$
4. x_3 today + Y' or x_4 in three days with $p = 0.9$ + \$$z$

Then, X, X', Y, and Y' are the differences between the utilities of the corresponding right-hand side (RHS) and the left-hand side (LHS) options.

$$X = u(RHS)_X - u(LHS)_X$$

Let us suppose that participants exhibit reference-dependent preferences such that $U(x|r) = u(x) + \mu(x)$. From having to do x tasks they get a (dis)utility of $u(x) = -d(x)$ plus a news-utility component that depends on whether they have to work more or less than their reference point:

$$\mu(x|r) = -\eta(d(r) - d(x)) \text{ if } x < r$$
$$\mu(x|r) = -\eta\lambda(d(x) - d(r) \text{ if } x \geq r$$

Since doing the tasks yield negative utility, having to do *more* tasks than the reference point is *negative* news. Also, having to do tasks today yields negative news utility compared to the reference point of no work today, but it also entails no tasks in three days, which has positive news utility compared to the expectation to work in three days. Let us assume that the participants set the fixed option in the price lists as their reference point. Assume that participants have present bias β but they are fully patient, so $\delta = 1$. In all of our batches, $x_1 \geq x_2, x_4$ and $x_3 \geq x_2, x_4$, to make sure that the LHS option is worse than the RHS option. $d(x)$ is the disutility of doing x tasks, and $d(x, p)$ is the disutility of a lottery where one has to do x task with probability p, and zero tasks otherwise. Let us further assume that $\eta = 1$ and the utility of money to be linear:

$$\begin{aligned}
X &= (1+\lambda)\beta d(x_1) - (1+\lambda)\beta d(x_2) \\
X' &= (1+\lambda)d(x_3) - 2\beta d(x_2) \\
Y &= (1+\lambda)\beta d(x_1) - (1+\lambda)\beta d(x_4, 0.9) + \$z \\
Y' &= (1+\lambda)d(x_3) - 2\beta d(x_4, 0.9) + \$z
\end{aligned}$$

Then,

$$\Delta\Delta = (\lambda - 1)\beta(d(x_4, 0.9) - d(x_2))$$

Therefore, the direction of $\Delta\Delta$ depends on the disutility difference between the excuse option and the no-excuse option. If the excuse option yields higher utility for most subjects than the no-excuse option, i.e. $d(x_4, 0.9) - d(x_2) < 0$, then reference dependence would predict a negative $\Delta\Delta$. In our first experiment, only Batch 2 met this requirement, but the results were similar in both batches. In the second experiment we made sure that the excuse option is always at least as good as the no-excuse option, so that if we find a positive $\Delta\Delta$, that means a strong excuse-driven present bias *despite* potentially reference-dependent preferences.

B.4 First experiment instructions

B.4.1 Session 0

HIT Description

This page describes the HIT, please read it closely.

Counting the number of 1's in small or large matrices

There are two types of matrices (see below for screenshots):

- **small matrices**: these are small matrices of 0's and 1's
- **large matrices**: these are larger matrices of 0's and 1's.

In all matrices, you have to count the number of 1's. See below for examples of matrices.

The HIT requires you to complete 6 matrices **correctly**.

You fail the matrix counting **and the HIT** if you give a wrong answer for more than 6 matrices.

Reading the description of a research study

After completing the HIT, you can participate in a research study. Before you decide if you want to participate in the study, we ask you to read a detailed description of the study and to complete a comprehension check.

> You do **not** have to participate in the study to complete the HIT successfully. If you complete the HIT, you will receive the HIT payment in 3-4 days.

Completing a quick survey

After deciding to participate in the study or not, we ask you some final questions about the HIT and the study.

Example of small matrix

0	0	1	0	0	1	0	0	1	0	0	0
0	0	1	0	0	1	0	0	1	0	1	0
0	1	0	0	0	1	0	1	1	0	1	0
1	0	0	1	0	1	0	1	1	1	0	0
0	1	0	1	1	1	0	0	0	1	0	1
1	1	0	1	0	0	0	1	1	1	1	0
0	1	1	0	1	1	1	1	0	0	0	1

Example of large matrix

0	1	1	1	1	0	1	0	0	1	1	0	0	1	0
1	0	1	0	0	0	1	1	1	0	0	0	1	1	1
1	0	1	1	1	1	0	0	1	1	0	1	1	0	1
0	0	0	0	1	0	1	1	1	1	1	0	1	0	0
0	0	1	0	0	0	0	0	1	1	0	1	0	0	1
1	0	1	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	1	1	1	0	0	1	0
1	1	0	1	0	0	1	0	0	0	1	1	1	0	1
0	1	0	0	0	0	1	0	1	0	0	1	1	1	0
0	0	0	0	0	1	0	0	0	1	0	0	1	0	0

Next

Matrix 1 of 6 (3 small matrices, 3 large matrices)

You can get at most 6 answers wrong.

Number of correct answers so far: 0/6

Number of wrong answers so far: 0/6

Count the cells containing 1

0	0	0	0	1	0	1	0	0	1	0	0
0	0	0	0	0	0	1	0	1	1	0	0
1	1	1	0	0	0	0	0	1	1	1	1
0	0	0	1	1	1	0	0	0	0	0	0
0	1	0	1	0	0	1	0	0	1	1	1
1	0	1	1	1	1	0	0	1	1	0	1
1	1	0	1	0	1	0	0	1	0	0	0

Number you counted:

Next

Matrix 4 of 6 (3 small matrices, 3 large matrices)

You can get at most 6 answers wrong.

Correct!

Number of correct answers so far: 3/6

Number of wrong answers so far: 0/6

Count the cells containing 1

1	1	1	1	1	0	1	0	1	1	1	0	0	1	1
1	0	1	1	0	0	1	0	0	1	1	0	0	1	1
1	1	0	0	0	0	1	0	0	1	0	0	0	0	0
1	1	1	0	0	1	0	0	1	0	0	0	1	0	1
0	1	1	0	0	1	0	0	0	1	1	0	0	1	0
1	0	0	0	1	0	0	1	0	0	0	1	1	1	1
0	0	0	1	0	1	1	0	1	1	0	0	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	1	1	0	0	0	1
1	1	1	1	0	1	0	1	0	0	0	1	0	0	1

Number you counted:

Next

10.14754/CEU.2022.08

Description of the Study following the HIT

The study you can join consists of two sessions:

1. the first session is **today** after completing the HIT;
2. the second session is **the day after tomorrow**.

You will have to use your MTurk ID to login to these sessions with links that you receive if you decide to participate.

Count 25 matrices correctly per session

In both sessions, you have to count **25 matrices correctly**. In one of the sessions, you will only have small matrices to count. In the other session, you will have to complete between 0 and 25 large matrices and the remaining matrices for that session will be small ones.

You fail the study if in either session 1 or session 2 you get more than 25 matrices wrong before getting 25 matrices right -- meaning you receive no completion bonus.

The number of large matrices to do is determined by your choices in session 1

At the beginning of session 1, we ask you to make several (~15) choices. These choices determine two things:

- By how much your completion bonus increases, which starts at $10.00
- How many of the 25 matrices on each day are small matrices, and how many are large matrices

You will face the choices as price lists. This is an example price list:

Screen shot of one price list

Add $3.00 to bonus; 22 large matrices today	Add $0.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.40 to bonus; 20 large matrices today

Comprehension questions and survey

There will be a comprehension check in session 1 and a quick survey (~1-2 minutes) at the end of session 2.

Payment

You receive your HIT payment whether or not you participate in the study. The payments below are in addition to the HIT fee. All study payments are made in 3-4 days – 1 or 2 days after the second session – as bonus payments to the HIT.

If you participate in the study and **complete** it, you will receive:

- a completion bonus for of $10.00 for session 2 plus the bonus determined from your choices (between $0.00 and $5.00)
- a session bonus of $1.00 for completing session 1.

If you participate in the study, but only complete the first session, you will receive only the session 1 bonus of $1.00.

If you participate in the study, but fail to complete session 1, you will receive no payment for the study. Therefore, **if you decide to participate in the study, make sure to come back and complete session 2.**

Study Purpose

This study is part of several research studies examining how people make decisions for work over time in the presence of uncertain outcomes and to measure how these decisions change.

Further Information

This study is conducted by Luca Drucker and Marc Kaufmann and financed by Central European University.

You can choose if you accept to participate in this study. If you accept to participate, you may still change your mind and leave the study at any time. In this case, you will not receive any bonus payments.

Participation in this experiment is not associated with any foreseeable risk or benefit.

Your answers will be collected confidentially and anonymously (the researchers will not be able to link decisions and participants' identity, beyond the MTurk worker ID provided).

In case the results of the study are published, there can be no references to your identity. Data anonymity is guaranteed.

If you have any questions or concerns regarding this study, please contact us at mkaufmann.research@gmail.com.

Next

Comprehension Check

Can you complete the HIT without participating in the study?

- ○ Yes
- ○ No

If you decide to participate in the study, but only complete today's session (make choices and complete the 25 matrices), how much do you receive (on top of the HIT participation fee)?

[ ] $

When is session 2?

- ○ Today
- ○ Tomorrow
- ○ The day after tomorrow
- ○ I am not sure

Next

10.14754/CEU.2022.08

Consent Form

Reminder: You do not need to participate in the study to complete the HIT.

Do you accept to participate in this study?

- ○ I decline
- ○ I accept

Next

For reference, the box below contains the description of the study provided earlier:

Next steps

You have decided to continue to the study.

- Bookmark the link to the first session: https://otree.trichotomy.xyz/room/first_room1/.
- You should shortly receive an email with the same link.
  - You will need to provide your MTurk Worker ID to login.
- Complete this HIT by clicking "Next" and completing a short one-page survey.
- Then start session 1 with the above link.

Next

10.14754/CEU.2022.08

Final Survey

Please answer the following questions.

Were the instructions of the study clear?

- ○ Yes
- ○ No

If not, what was unclear about the instructions?

On a scale from 0 (much more unpleasant) to 10 (much more pleasant), how does the task of counting small matrices compare to other tasks you usually do on MTurk?

On a scale from 0 (much more unpleasant) to 10 (much more pleasant), how does the task ofcounting large matrices compare to other tasks you usually do on MTurk?

How much time (in hours) do you usually spend on MTurk HITs per week?

How much money (in $) do you usually make from MTurk HITs per week?

In percent (0% meaning not at all, to 100% meaning you are absolutely sure), how likely it is that you will complete session 1?

In percent (0% meaning not at all, to 100% meaning you are absolutely sure), how likely it is that you will complete both session 1 and session 2?

How can we make it easier for participants to participate in both sessions given they are one or several days apart?

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 means not willing at all, 10 means very willing:

Next

10.14754/CEU.2022.08

Thanks

Thanks for participating. You are done and the HIT will be completed by clicking 'Next'.

Here is again the link to the first session: https://otree.trichotomy.xyz/room/first_room1/.

You will receive your payment within 3-4 days.

Next

B.4.2 Session 1

Welcome

Before you do today's tasks, we will show you 15 pages. On each page, you will make multiple choices between two options (see screenshot below).

After you have made your choices, we randomly choose one page, and one choice from that page. That choice will count and it will determine:

- your extra bonus payment (on top of the $10.00 for completing session 2),
- how many of the 25 matrices that you have to complete will be large matrices rather than small ones, and in what session you will have to complete these large matrices,
- the probability with which you will have to do these large matrices rather than small ones (see example below).

Remember that you only receive your extra bonus payment if you complete all aspects of the study, so both the current session and session 2.

Example

Example 1: Suppose the choice you made and that was picked was "Add $2.00 to bonus; 20 large matrices today". Then

- you will receive $2.00 + $10.00 (session 2) + $1.00 (session 1) if you complete the study,
- you will do 20 large matrices today; the remaining 5 matrices will be small ones,
- you will do no large matrices the day after tomorrow; so you will do 25 small ones the day after tomorrow.

Example 2: Suppose the choice you made and that was picked was "Add $3.00 to bonus; 22 large matrices (80%) today". Then

- you will receive $3.00 + $10.00 (session 2) + $1.00 (session 1) if you complete the study,
- there is an 80% chance that you will have to do 22 large matrices and 3 small matrices today
- and there is **a 20% chance that you do not have to do any large matrices** today, in which case you do 25 small matrices today.
- Whether you have to do large matrices will be determined right before you start doing any tasks.
- You will do 0 large matrices and 25 small ones the day after tomorrow.

10.14754/CEU.2022.08

Screenshot of a Page

Add $3.00 to bonus; 22 large matrices today	Add $0.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.40 to bonus; 20 large matrices today

Check questions

When will you receive all payments?

- Today
- Right at the end of session 2
- Within 1 or 2 days after completing session 2

Suppose you chose 'Add $2.00 to bonus; 20 large matrices today'. Will you receive the $2.00 if you do not complete the whole study?

- Yes
- No

Which of the following statements is true if the following option is chosen: 'Add $3.00 to bonus; 22 large matrices (80%) today'?

- You will receive $3.00 bonus even if you don't complete the whole study
- There is a 20% chance that you will not have to do any large matrices today and your bonus from completing the study is $3.00 higher
- There is a 20% chance that you will not have to do any large matrices today and your bonus from completing the study won't be higher

Suppose you chose 'Add $2.00 to bonus; 20 large matrices today'. Which of the following is true?

- This means you have to do 45 matrices in total today: 20 large matrices and 25 small matrices.
- This means you have to do 25 matrices in total today: 20 large matrices and 5 small matrices
- This means you have to do 20 matrices in total today: 20 large matrices and no small matrices

To continue to the choices, click on the Next button.

Next

Choice 1

Notes

- The options in the left column stay the same, while choices in the right column have increasing payments
- If you complete the study, you will always receive the extra payment, even when it is uncertain whether you have to do large matrices.

For each row, pick the choice that you prefer.

○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $0.00 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $0.20 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $0.40 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $0.60 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $0.80 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $1.00 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $1.20 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $1.40 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $1.60 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $1.80 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $2.00 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $2.20 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $2.40 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $2.60 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $2.80 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $3.00 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $3.20 to bonus; 5 large matrices in two days
○ Add $2.40 to bonus; 19 large matrices (90%) in two days	○ Add $3.40 to bonus; 5 large matrices in two days

Next

Random draw of choice to implement

Now we randomly choose one of the 15 pages on which you chose, and we pick a random choice on that page. This choice will determine

- your extra bonus payment,
- when you will have to do large matrices,
- how many large matrices you will have to do.

Next

Results

Page 12 was chosen, and the choice was between

- Add $0.20 to bonus; 23 large matrices today
- Add $0.20 to bonus; 19 large matrices (90%) in two days

You chose "Add $0.20 to bonus; 23 large matrices today".

Next

Question about the preceding choices

When making the choices, how did you decide which option to choose?

When will you receive all payments?

How much bonus will you receive if you only complete today's session and don't come back for session 2?

Was it clear that no matter what you choose, you will have 25 matrices to do, and that your choice only affects how many large and how many small matrices you will do?

Next

10.14754/CEU.2022.08

Instructions

You have to do 23 large matrices; the remaining matrices will be small matrices.

Only a correctly solved matrix counts as a completed task. If you enter a wrong number, you will be shown another matrix to try again. **If you get more than 25 matrices wrong, then you will automatically drop out of the study without bonus or completion payments.**

Click on the next button to continue to the tasks.

Next

End of this session's tasks

Congratulations, you are finished with this session's 25 tasks.

Use this link to come back for the next session:https://otree.trichotomy.xyz/room/first_room2/. Make sure to bookmark it or store it where you can find it.

Reminder: Make sure to set yourself a reminder to come back after 6am Pacific Coast Time the day after tomorrow (or follow the link and check how much time is left until it opens). If you don't come back, you will not receive the $10.00 completion bonus or the extra payments.

Next

10.14754/CEU.2022.08

B.4.3 Session 2

Without uncertainty:

Instructions

You have to do 5 large matrices; the remaining matrices will be small matrices.

Only a correctly solved matrix counts as a completed task. If you enter a wrong number, you will be shown another matrix to try again. **If you get more than 25 matrices wrong, then you will automatically drop out of the study without bonus or completion payments.**

Click on the next button to continue to the tasks.

Next

With uncertainty:

Determining whether you have to do extra tasks

We now determine randomly whether you have to do the 5 large matrices assigned for today, which happens with a probability of 90.

Click on the next button to continue.

Next

Resolving uncertainty:

Instructions

You don't have to do the 5 large matrices! Thus you do 25 small matrices.

Only a correctly solved matrix counts as a completed task. If you enter a wrong number, you will be shown another matrix to try again. **If you get more than 25 matrices wrong, then you will automatically drop out of the study without bonus or completion payments.**

Click on the next button to continue to the tasks.

Next

10.14754/CEU.2022.08

B.5 Second experiment instructions

Welcome

You are about to start the tutorial of the study. During the tutorial you will:

- read the description of the full study that you can sign up for
- do practice tasks to familiarize yourself with the tasks from the study
- make practice choices for extra tasks like those asked in the study
- answer comprehension questions to check you understand the study
- decide if you want to participate in the study

The payment for the tutorial does not depend on whether you participate in the study or not. You will receive the tutorial payment of €4.00 whether you sign up for the study or not:

- You receive the tutorial payment if you complete the tutorial, which requires getting enough tasks right.
- If you do not participate in the study, you receive your payment within a week
- If you participate in the study, you receive all your payments in one go in the week after the end of the study.

Continue

DigiCash or Paypal required

We can only pay you for your participation via DigiCash (preferred), or via Paypal with an electronic signature. For these reasons, you need to activate DigiCash on your Luxembourgish bank account; or open a Paypal account to participate in this study.

I understand that I need Digicash or open a Paypal account to be paid for this study ☐ Continue

10.14754/CEU.2022.08

Study Description

The study consists of:

- one session today after the tutorial
- one session in two days
- the sessions can be completed any time of the day in GMT+2 timezones (Berlin/Luxembourg/Vienna)
- a final survey

You will receive reminder email for the second session.

Default Tasks

In both sessions, you have to do 20 matrix counting tasks:

- Matrix tasks: counting the number of 1's in a matrix of 1's, 10's, 11's, and 0's

By default, you have to count small matrices. However, in the first session, you will make decisions about when and how many **large matrix tasks instead of small matrix tasks** you will do for bonus payments. We randomly pick one of the choices as the choice that counts:

- This choice is picked from all the choices you made, so all choices can become the choice that counts.
- The bonus of that choice will be added to the completion bonus, if you complete both sessions that week.
- The choice that counts determines in which session you will do large matrices, and how many of the 20 tasks in that session are going to be large matrices rather than small.
- You will have to do large matrices only in one of the sessions, either in session 1 or session 2, and in each session the number of small and large matrices always adds up to 20.

Payment

You receive all your payments in the week after the end of the study, even if you drop out earlier.

You get payment for completing the following:

- Completing today's tutorial: €4.00
- Completing both sessions, including all small and large tasks: €15.00
- Bonus depending on your choice

You do not receive partial payment or the bonus if you complete only one session or only some but not all of the tasks of a particular session! If you miss a session, you cannot do it on another day, no matter the reason for missing it.

Continue

Description: small Matrix Tasks

You will now do 3 small Matrix tasks.

Matrix Task Description

Each page will look like the screenshot below.

- You have to count the number of cells that contain exactly the number 1
- Cells containing the number 11 or 10 do not count
- If you get more tasks wrong than you have to get right, you fail the tasks and thus the tutorial
- If you get a matrix wrong, you will be given a new one.

Screenshot of page

Count only cells with exactly "1" in it

In this toy 2 by 2 matrix, only 1 cell contains exactly the number 1. The cells containing 10 and 11 do not count:

0	11
1	10

Example of a small matrix

Here is an example of a small matrix:

Count the cells with 1's in them

You completed 0 out of 3 tasks (0 wrong guesses out of at most 3)

If you get more than 3 wrong guesses, you drop out of the study.

00	1	0	1	11	1	1	0	0	1	1	1
00	1	0	0	11	1	00	11	11	00	11	0
1	1	10	10	11	0	10	00	10	00	1	0
0	0	0	1	1	0	1	0	1	1	0	10
11	0	0	0	11	0	0	11	0	0	1	0
0	0	0	11	1	1	1	10	1	1	10	0
11	1	00	0	10	0	10	11	1	11	1	10

How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

Submit

10.14754/CEU.2022.08

Example of a large matrix

Here is an example of a large matrix:

Count the cells with 1's in them

You completed 0 out of 3 tasks (0 wrong guesses out of at most 3)

If you get more than 3 wrong guesses, you drop out of the study.

1	11	1	00	0	1	0	0	11	1	0	1	11	0	0
0	00	1	1	0	0	0	1	0	1	1	1	0	0	1
11	00	0	0	1	1	10	1	0	1	1	0	0	1	0
00	1	1	0	10	1	1	0	0	1	1	1	1	10	00
00	0	0	0	1	0	0	1	1	0	0	11	0	0	1
1	10	1	11	1	10	1	11	1	00	1	10	0	11	1
1	1	0	11	1	1	1	11	10	1	11	0	1	00	1
1	00	0	11	00	1	0	1	0	11	0	10	11	0	0
00	1	1	0	00	1	0	0	10	1	00	0	1	10	00
0	10	10	1	0	00	1	11	1	1	11	0	1	00	11

How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

Submit

Start tasks

Count the cells with 1's in them

You completed 0 out of 3 tasks (0 wrong guesses out of at most 3)

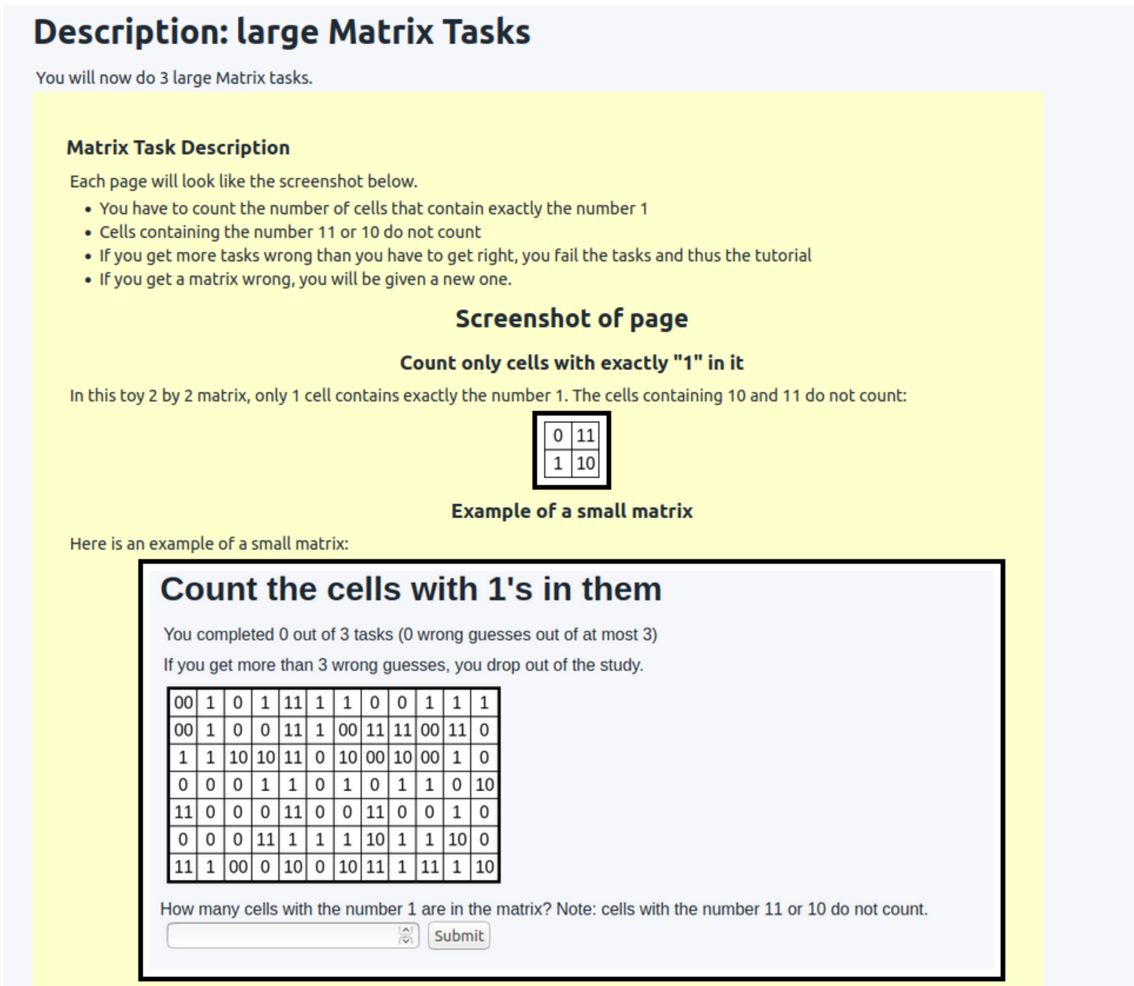
If you get more than 3 wrong guesses, you drop out of the study.

00	0	0	0	0	1	0	1	0	0	1	00
1	1	0	0	0	10	1	0	00	1	0	1
1	0	1	0	1	1	0	0	1	0	0	0
0	1	0	1	1	1	1	1	0	1	1	00
1	11	1	0	0	1	1	1	0	11	0	1
1	1	1	1	1	1	0	00	0	10	0	1
1	0	1	10	00	1	0	1	1	1	0	0

How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

Submit

10.14754/CEU.2022.08



Description: large Matrix Tasks

You will now do 3 large Matrix tasks.

Matrix Task Description

Each page will look like the screenshot below.

- You have to count the number of cells that contain exactly the number 1
- Cells containing the number 11 or 10 do not count
- If you get more tasks wrong than you have to get right, you fail the tasks and thus the tutorial
- If you get a matrix wrong, you will be given a new one.

Screenshot of page

Count only cells with exactly "1" in it

In this toy 2 by 2 matrix, only 1 cell contains exactly the number 1. The cells containing 10 and 11 do not count:

0	11
1	10

Example of a small matrix

Here is an example of a small matrix:

Count the cells with 1's in them

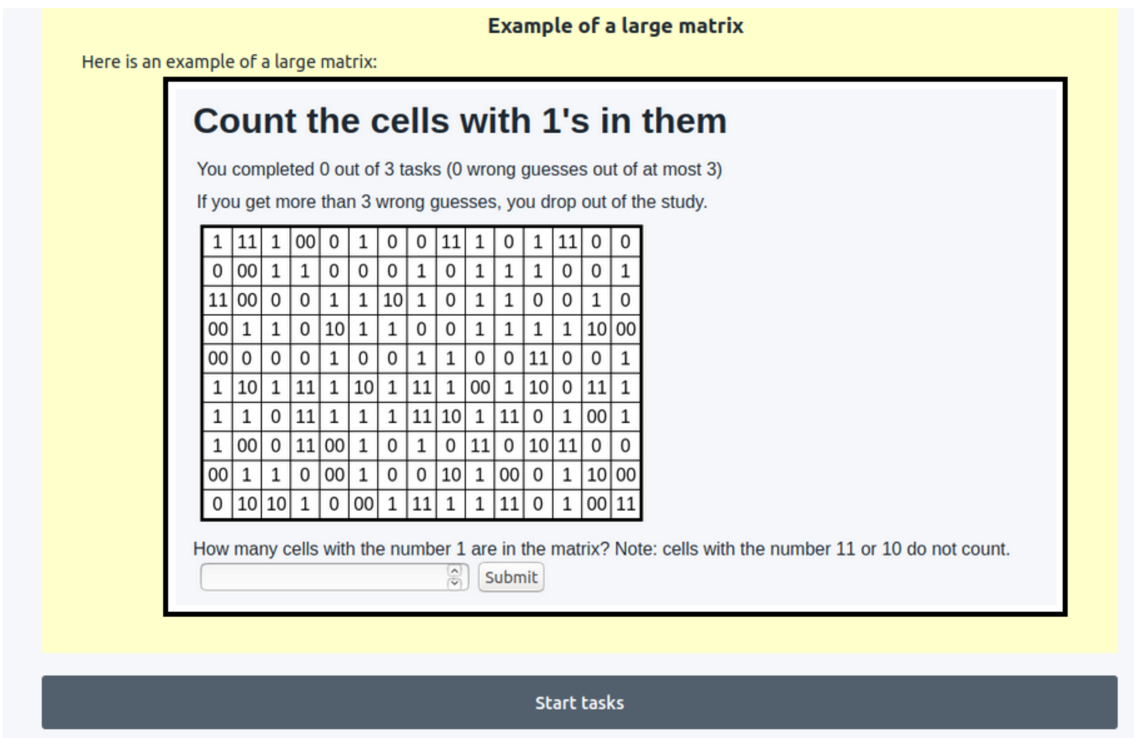
You completed 0 out of 3 tasks (0 wrong guesses out of at most 3)

If you get more than 3 wrong guesses, you drop out of the study.

00	1	0	1	11	1	1	0	0	1	1	1
00	1	0	0	11	1	00	11	11	00	11	0
1	1	10	10	11	0	10	00	10	00	1	0
0	0	0	1	1	0	1	0	1	1	0	10
11	0	0	0	11	0	0	11	0	0	1	0
0	0	0	11	1	1	1	10	1	1	10	0
11	1	00	0	10	0	10	11	1	11	1	10

How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

Submit



Example of a large matrix

Here is an example of a large matrix:

Count the cells with 1's in them

You completed 0 out of 3 tasks (0 wrong guesses out of at most 3)

If you get more than 3 wrong guesses, you drop out of the study.

1	11	1	00	0	1	0	0	11	1	0	1	11	0	0
0	00	1	1	0	0	0	1	0	1	1	1	0	0	1
11	00	0	0	1	1	10	1	0	1	1	0	0	1	0
00	1	1	0	10	1	1	0	0	1	1	1	1	10	00
00	0	0	0	1	0	0	1	1	0	0	11	0	0	1
1	10	1	11	1	10	1	11	1	00	1	10	0	11	1
1	1	0	11	1	1	1	11	10	1	11	0	1	00	1
1	00	0	11	00	1	0	1	0	11	0	10	11	0	0
00	1	1	0	00	1	0	0	10	1	00	0	1	10	00
0	10	10	1	0	00	1	11	1	1	11	0	1	00	11

How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

Submit

Start tasks

10.14754/CEU.2022.08

B.5.1 Session 1 each week

Decisions for doing large matrices instead of small matrices

In the first session, you face 20 pages of choices about doing large matrices instead of small ones in one of the sessions. Different choices involve different bonuses. We pick one of the pages at random, and one choice from that page at random as **the choice that counts**. You will face the following types of options to choose from:

Add €1.00 to bonus; 10 large matrices in two days

You have to do 10 large matrix tasks in 2 days, and you receive €1.00 on top of your completion bonus after completing the study.

Add €1.50 to bonus; 10 large matrices (90%) in two days

There is a 90% chance that you have to do 10 large matrix tasks in 2 days; and a 10% chance that instead you do 10 small matrix tasks that day. In both cases you receive €1.50 on top of your completion bonus. You will learn whether you have to do the large or small tasks in 2 days.

You will now see two such **practice decision pages** to illustrate the how the choice that counts is determined.

Go to practice choices

Practice Only!

Decision Page : large instead of small matrices

Please choose for each row. You can use the slider to choose the payment for which you pick the left to fill the choices below or click on your preferred choice in each row directly.

Reminder: Choosing large matrices means that these large matrices **replace** small matrices, they are not in addition. Thus you always do 20 tasks in each session.

"I choose the left side when the payment is at least €0.00."

○ Add €0.00 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €0.20 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €0.40 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €0.60 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €0.80 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €1.00 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €1.20 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €1.40 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €1.60 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €1.80 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €2.00 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €2.20 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €2.40 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €2.60 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days
○ Add €2.80 to bonus; 15 large matrices today	○ Add €1.00 to bonus; 10 large matrices (90%) in two days

10.14754/CEU.2022.08

Determining the choice that counts

In the study, we will randomly pick one of the pages of choices you made, and you have to do the tasks as determined by that choice.

- Then we randomly pick one choice from that page.
- That choice is the one that counts: you will receive the bonus corresponding to that choice **if you complete all parts of the study!**

Let us illustrate this process on the next page based on the two pages of choices you just made.

Continue

Choice that counts

In the 5th choice of the second page, you chose

- **Add €1.00 to bonus; 10 large matrices today** instead of
- **Add €0.80 to bonus; 20 large matrices today**

Remember: If a choice is about '10 large matrices (90%)', then:

- There is a 90% chance that you have to do 10 small matrices and 10 large matrices, and a 10% chance that you do only small matrices - i.e. 20 small matrices, no large matrices.
- This is determined on the day the tasks are potentially due
- You will receive the additional bonus independent of whether you have to do large matrices or not.

Continue

Comprehension Test

In order to participate in the study, you have to pass this comprehension test. Refer to the study description below the questions for help.

Beyond the bonus for the tutorial, how much do you receive if you participate in the study and complete only one of the two sessions?

- You would receive nothing - only full completion leads to payment.
- You would receive half the payment.
- You would receive the bonus payment for any large matrices.

How many tasks are there in every session in total?

20

Suppose that the choice that counts is such that you have to do 10 large matrices in session 1. This means that in session 1...

- ... you would have to do 20 small matrices and 10 large matrices in session 1.
- ... you would have to do no small matrices and 10 large matrices.
- ... you would have to do 10 small matrices and 10 large matrices.

Suppose that the choice that counts is such that you have to do 10 large matrices in session 1. This means that in session 2...

- You do not have enough information to know how many small and large matrices you may have to do in session 2.
- ... you would have to do 20 small matrices and no large matrices.
- ... you would have to do 10 small matrices and 10 large matrices.

Which of the following statements is true if the following option is chosen: 'Add $3.00 to bonus; 15 large matrices (90%) today'?

- You will receive $3.00 bonus even if you don't complete the whole study
- There is a 10% chance that you will not have to do any large matrices today, so you do 20 small matrices, and your bonus from completing the study is $3.00 higher
- There is a 10% chance that you will not have to do any large matrices today and your bonus from completing the study won't be higher

Continue

Consent Form

Do you agree to participate in the full study? The description is repeated below the form.

(Note: You do not need to participate in the study to receive the bonus for the tutorial.)

○ I do agree to participate
○ I do not agree to participate

Continue

Study Description

The study consists of:

- one session today after the tutorial
- one session in two days
- the sessions can be completed any time of the day in GMT+2 timezones (Berlin/Luxembourg/Vienna)
- a final survey

You will receive reminder email for the second session.

Default Tasks

In both sessions, you have to do 20 matrix counting tasks:

- Matrix tasks: counting the number of 1's in a matrix of 1's, 10's, 11's, and 0's

By default, you have to count small matrices. However, in the first session, you will make decisions about when and how many **large matrix tasks instead of small matrix tasks** you will do for bonus payments. We randomly pick one of the choices as the choice that counts:

- This choice is picked from all the choices you made, so all choices can become the choice that counts.
- The bonus of that choice will be added to the completion bonus, if you complete both sessions that week.
- The choice that counts determines in which session you will do large matrices, and how many of the 20 tasks in that session are going to be large matrices rather than small.
- You will have to do large matrices only in one of the sessions, either in session 1 or session 2, and in each session the number of small and large matrices always adds up to 20.

Payment

You receive all your payments in the week after the end of the study, even if you drop out earlier.

You get payment for completing the following:

- Completing today's tutorial: €4.00
- Completing both sessions, including all small and large tasks: €15.00
- Bonus depending on your choice

You do not receive partial payment or the bonus if you complete only one session or only some but not all of the tasks of a particular session! If you miss a session, you cannot do it on another day, no matter the reason for missing it.

20 decision pages coming up

Next you will see 20 pages of decisions for extra work for this week.

Continue

Decision Page : large instead of small matrices (1 of 20)

Please choose for each row. You can use the slider to choose the payment for which you pick the left to fill the choices below or click on your preferred choice in each row directly.

Reminder: Choosing large matrices means that these large matrices **replace** small matrices, they are not in addition. Thus you always do 20 tasks in each session.

"I choose the left side when the payment is at least €0.00."

◉ Add €0.00 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €0.20 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €0.40 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €0.60 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €0.80 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €1.00 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €1.20 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €1.40 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €1.60 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €1.80 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €2.00 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €2.20 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €2.40 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €2.60 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €2.80 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €3.00 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days
◉ Add €3.20 to bonus; 12 large matrices in two days	○ Add €0.20 to bonus; 8 large matrices in two days

Choice that counts for week 1

The choice that counts is: Add €3.60 to bonus; 18 large matrices (90%) today

Continue

Description: small Matrix Tasks

You will now do 2 small small Matrix tasks.

Show Task Description

Start tasks

Description: large Matrix Tasks

You will now do 18 large large Matrix tasks.

Show Task Description

Start tasks

Count the cells with 1's in them

You completed 0 out of 18 tasks (0 wrong guesses out of at most 18)

If you get more than 18 wrong guesses, you drop out of the study.

1	10	00	1	00	11	00	10	1	0	11	1	1	00	1
1	1	1	11	0	0	1	0	0	0	10	1	1	1	1
0	0	0	1	0	0	10	1	0	00	1	0	1	0	1
0	1	1	00	0	0	1	1	1	1	0	0	1	1	1
11	1	0	1	0	0	11	0	1	1	0	1	1	1	1
1	1	1	0	1	1	1	10	1	0	0	0	10	00	11
00	10	1	10	1	1	10	0	0	10	1	10	1	00	1
1	1	0	0	1	0	10	0	0	0	0	10	0	1	1
1	10	1	10	0	0	1	1	1	0	0	11	10	1	10
1	1	0	0	0	1	0	1	1	1	11	10	1	0	1

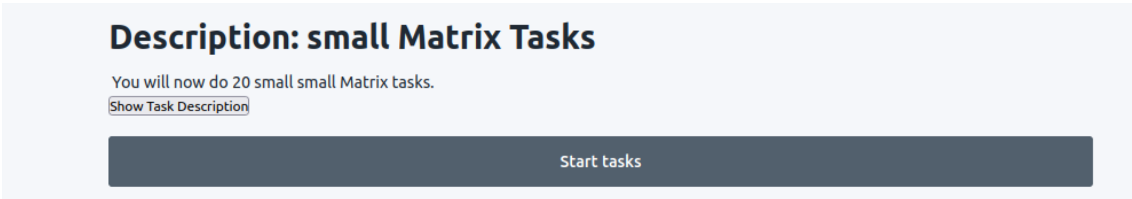
How many cells with the number 1 are in the matrix? Note: cells with the number 11 or 10 do not count.

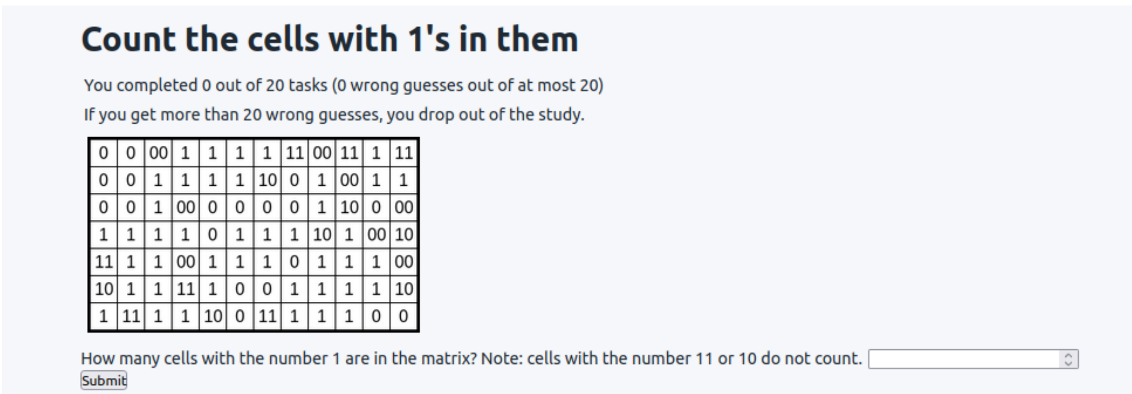
Submit

You are done with session 1

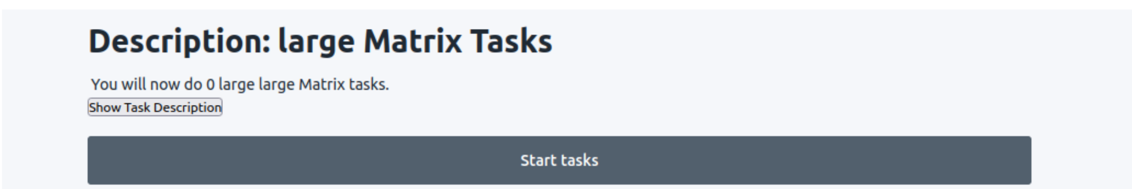
Continue

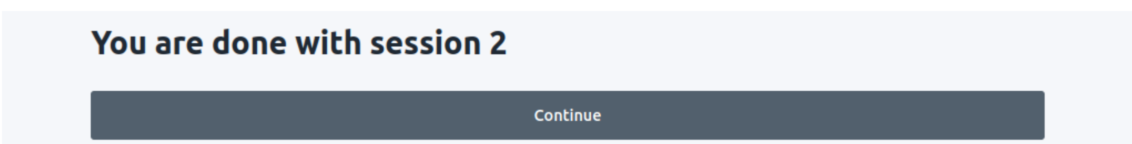
B.5.2 Session 2 each week



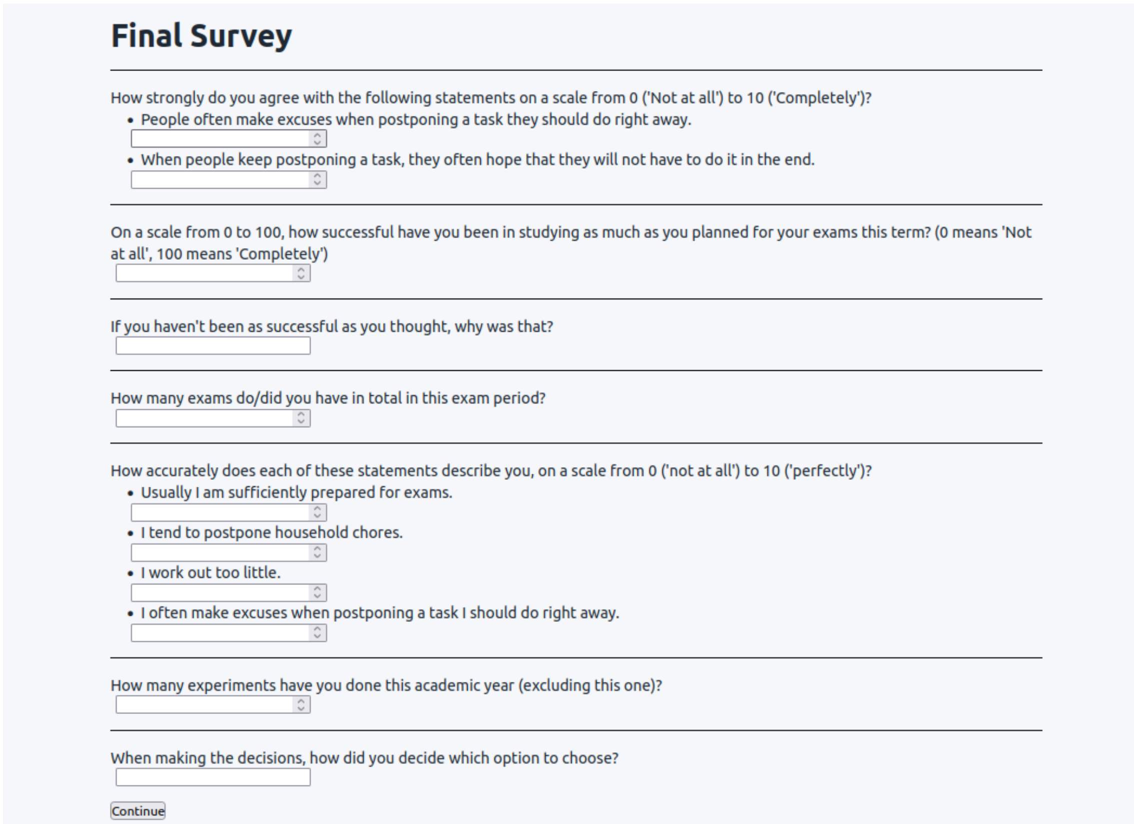


0	0	00	1	1	1	1	11	00	11	1	11
0	0	1	1	1	1	10	0	1	00	1	1
0	0	1	00	0	0	0	0	1	10	0	00
1	1	1	1	0	1	1	1	10	1	00	10
11	1	1	00	1	1	1	0	1	1	1	00
10	1	1	11	1	0	0	1	1	1	1	10
1	11	1	1	10	0	11	1	1	1	0	0





10.14754/CEU.2022.08



Final Survey

How strongly do you agree with the following statements on a scale from 0 ('Not at all') to 10 ('Completely')?

- People often make excuses when postponing a task they should do right away.
- When people keep postponing a task, they often hope that they will not have to do it in the end.

On a scale from 0 to 100, how successful have you been in studying as much as you planned for your exams this term? (0 means 'Not at all', 100 means 'Completely')

If you haven't been as successful as you thought, why was that?

How many exams do/did you have in total in this exam period?

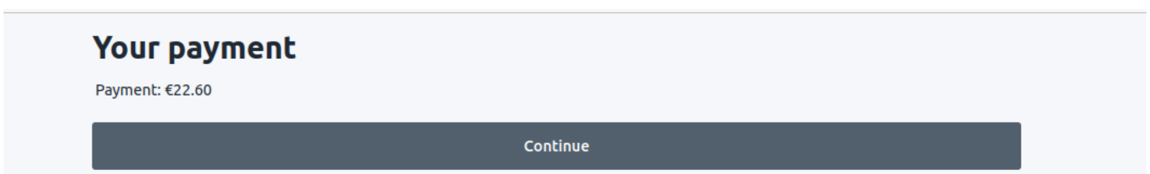
How accurately does each of these statements describe you, on a scale from 0 ('not at all') to 10 ('perfectly')?

- Usually I am sufficiently prepared for exams.
- I tend to postpone household chores.
- I work out too little.
- I often make excuses when postponing a task I should do right away.

How many experiments have you done this academic year (excluding this one)?

When making the decisions, how did you decide which option to choose?

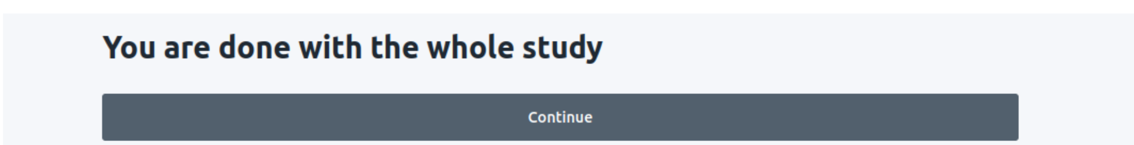
Continue



Your payment

Payment: €22.60

Continue



You are done with the whole study

Continue

10.14754/CEU.2022.08

C Appendix for Chapter 3

C.0.1 Descriptive tables

	Socio-economic status		
6th grade aspirations in 3 categories	Low SES	High SES	Total
At most vocational	0.0298 40175	0.0213 11056	0.0280 51231
High school diploma or post-secondary non-tertiary	0.0104 82594	0.0063 80184	0.0084 162778
College or university	0.0036 41555	0.0014 203903	0.0018 245458
Total	0.0135 164324	0.0035 295143	0.0071 459467

Table C.1: Observations and share of repeaters in each aspiration-SES cell

Note: Each cell contains the share of repeaters in the cell and the number of observations in the cell. The sample for this table contains children who are in primary school in 6^{th} grade and who did not repeat the 6^{th} grade.

	Socio-economic status		
Repeated 7th grade	Low	High	Total
No	0.155	0.135	0.142
Yes	-0.154	0.070	-0.083
Total	0.151	0.135	0.141

Table C.2: Change in the educational aspirations in years by parental education

Note: Each cell shows the average change in the aspired years of education between 6^{th} grade and 8^{th} grade in the cell.

10.14754/CEU.2022.08

C.0.2 Retention probabilities

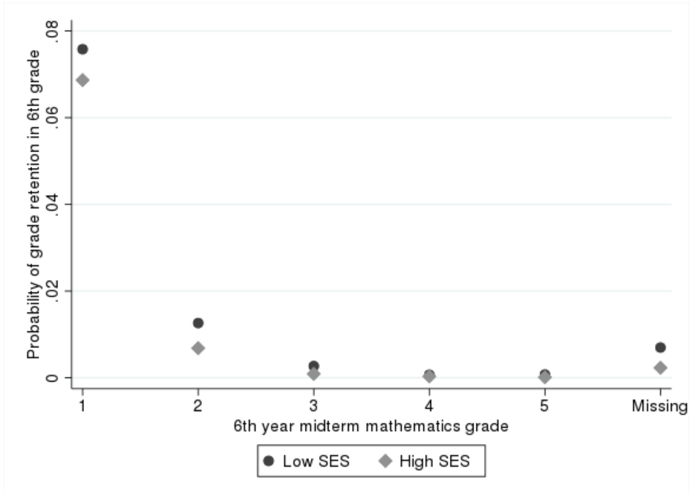


Figure C.1: Probability of repeating 6th grade by 6th year midterm mathematics grade

	1: Low-SES non-rep	1: Low-SES rep	1: High-SES non-rep	1: High-SES rep	2: Low-SES non-rep	2:Low-SES rep	2: High-SES non-rep	2:High-SES rep
6th grade mathematics test	1294.82	1308.36	1357.22	1315.58	1341.54	1309.53	1386.18	1373.29
score	(150.55)	(142.96)	(149.16)	(113.89)	(150.24)	(139.27)	(143.01)	(164.63)
6th grade reading test score	1283.34	1272.77	1359.71	1282.60	1332.81	1286.81	1392.26	1345.19
	(153.47)	(135.00)	(165.11)	(158.66)	(152.78)	(141.26)	(152.78)	(147.82)
6th year midterm literature	2.40	2.12	2.63	2.18	2.90	2.43	3.17	2.59
grade	(0.80)	(0.73)	(0.94)	(0.93)	(0.80)	(0.69)	(0.81)	(0.70)
6th year midterm Hungarian	2.23	1.99	2.44	2.13	2.71	2.32	2.91	2.56
grammar grade	(0.76)	(0.65)	(0.83)	(0.81)	(0.75)	(0.66)	(0.77)	(0.71)
Parents lived together in 6th	0.72	0.76	0.65	0.50	0.75	0.64	0.70	0.45
grade	(0.45)	(0.43)	(0.48)	(0.51)	(0.43)	(0.48)	(0.46)	(0.50)
Parents separated from 6th to	0.06	0.12	0.06	0.18	0.05	0.10	0.06	0.10
8th grade	(0.24)	(0.32)	(0.23)	(0.39)	(0.23)	(0.30)	(0.24)	(0.31)
Mother does not work in 6th	0.51	0.51	0.25	0.21	0.44	0.46	0.20	0.24
grade	(0.50)	(0.50)	(0.43)	(0.41)	(0.50)	(0.50)	(0.40)	(0.43)
Mother has a permanent job in	0.35	0.36	0.64	0.66	0.43	0.41	0.68	0.62
6th grade	(0.48)	(0.48)	(0.48)	(0.48)	(0.49)	(0.49)	(0.47)	(0.49)
Father does not work in 6th	0.24	0.23	0.10	0.18	0.19	0.24	0.09	0.11
grade	(0.43)	(0.43)	(0.30)	(0.39)	(0.39)	(0.43)	(0.29)	(0.32)
Father has a permanent job in	0.53	0.57	0.69	0.63	0.58	0.53	0.71	0.69
6th grade	(0.50)	(0.50)	(0.46)	(0.49)	(0.49)	(0.50)	(0.46)	(0.47)
Mother stopped working between	0.08	0.06	0.07	0.13	0.08	0.09	0.05	0.06
6th and 8th grades	(0.27)	(0.25)	(0.25)	(0.34)	(0.27)	(0.29)	(0.22)	(0.24)
Father stopped working between	0.07	0.09	0.04	0.05	0.07	0.06	0.04	0.03
6th and 8th grades	(0.26)	(0.28)	(0.21)	(0.23)	(0.26)	(0.25)	(0.20)	(0.17)
Number of GP visits in 7th	5.78	7.06	5.07	4.16	5.42	7.75	4.60	7.28
grade	(5.15)	(6.20)	(4.89)	(3.08)	(5.10)	(6.74)	(4.38)	(7.29)
Days spent in hospital in 7th	0.35	0.41	0.46	0.24	0.30	0.65	0.32	1.50
grade	(2.72)	(1.36)	(2.19)	(0.85)	(1.97)	(4.65)	(1.79)	(8.90)
Observations	1418	94	592	38	13720	296	9090	96

Table C.3: Characteristics of repeaters by 6^{th} year mathematics grades and SES: failing (1) and just passing (2) mathematics

10.14754/CEU.2022.08

	3-5: Low-SES non-rep	3-5: Low-SES rep	3-5: High-SES non-rep	3-5: High-SES rep
6th grade mathematics test	1481.94	1353.92	1580.84	1477.61
score	(164.12)	(163.08)	(166.98)	(162.14)
6th grade reading test score	1469.79	1345.19	1578.84	1464.13
	(165.77)	(169.70)	(166.89)	(187.27)
6th year midterm literature	3.92	2.98	4.39	3.52
grade	(0.87)	(0.94)	(0.73)	(1.07)
6th year midterm Hungarian	3.72	2.95	4.17	3.49
grammar grade	(0.87)	(0.86)	(0.80)	(1.05)
Parents lived together in 6th	0.78	0.69	0.79	0.55
grade	(0.41)	(0.47)	(0.40)	(0.50)
Parents separated from 6th to	0.05	0.13	0.04	0.12
8th grade	(0.22)	(0.33)	(0.20)	(0.33)
Mother does not work in 6th	0.34	0.49	0.14	0.23
grade	(0.47)	(0.50)	(0.35)	(0.42)
Mother has a permanent job in	0.52	0.43	0.75	0.61
6th grade	(0.50)	(0.50)	(0.44)	(0.49)
Father does not work in 6th	0.14	0.24	0.06	0.11
grade	(0.34)	(0.43)	(0.24)	(0.31)
Father has a permanent job in	0.66	0.53	0.72	0.62
6th grade	(0.47)	(0.50)	(0.45)	(0.49)
Mother stopped working between	0.07	0.10	0.04	0.10
6th and 8th grades	(0.25)	(0.30)	(0.19)	(0.30)
Father stopped working between	0.05	0.10	0.03	0.06
6th and 8th grades	(0.22)	(0.30)	(0.16)	(0.24)
Number of GP visits in 7th	4.81	7.51	3.99	6.16
grade	(4.69)	(7.23)	(4.03)	(7.43)
Days spent in hospital in 7th	0.27	1.36	0.24	2.20
grade	(1.64)	(6.89)	(1.69)	(11.02)
Observations	39674	103	102489	82

Table C.4: Characteristics of repeaters by 6^{th} year mathematics grades and SES: grades 3-5

C.0.3 Robustness checks for main regressions

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.378***	0.255***	0.153***	0.156***
	(0.011)	(0.013)	(0.013)	(0.012)	(0.012)
Medium SES		-0.179***	-0.104***	-0.0823***	-0.0688***
		(0.0024)	(0.0017)	(0.0014)	(0.0013)
High SES		-0.209***	-0.111***	-0.0819***	-0.0632***
		(0.0025)	(0.0016)	(0.0013)	(0.0013)
Repeated × Medium SES		-0.123***	-0.109***	-0.113***	-0.111***
		(0.025)	(0.023)	(0.022)	(0.022)
Repeated × High SES		-0.247***	-0.193***	-0.179***	-0.183***
		(0.026)	(0.023)	(0.023)	(0.023)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.376***	-0.289***	-0.275***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.430***	-0.294***	-0.273***
			(0.0034)	(0.0031)	(0.0030)
Constant	0.0991***	0.223***	0.515***	0.878***	0.813***
	(0.0019)	(0.0026)	(0.0036)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.

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

Table C.5: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma. Medium SES means at least one parent has a high school diploma. High SES means at least one parent has a tertiary degree. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

C.0.3.1 Using three categories of SES

10.14754/CEU.2022.08

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.215***	-0.0768***	0.000326	-0.000685
	(0.0072)	(0.0053)	(0.0057)	(0.0055)	(0.0056)
Medium SES		0.318***	0.167***	0.131***	0.115***
		(0.0025)	(0.0021)	(0.0019)	(0.0018)
High SES		0.581***	0.316***	0.251***	0.212***
		(0.0029)	(0.0027)	(0.0025)	(0.0021)
Repeated × Medium SES		-0.204***	-0.131***	-0.0931***	-0.0917***
		(0.017)	(0.017)	(0.017)	(0.017)
Repeated × High SES		-0.156***	-0.110***	-0.0564**	-0.0468*
		(0.032)	(0.027)	(0.025)	(0.024)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.146***	0.0496***	0.0412***
post-secondary non-tertiary			(0.0019)	(0.0018)	(0.0019)
College or university			0.620***	0.387***	0.365***
			(0.0023)	(0.0026)	(0.0026)
Constant	0.541***	0.254***	0.0178***	-0.505***	-0.417***
	(0.0043)	(0.0021)	(0.0019)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.

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

Table C.6: Probability of 8^{th} grade aspirations being tertiary education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. Low SES means both parents have less than a high school diploma. Medium SES means at least one parent has a high school diploma. High SES means at least one parent has a tertiary degree. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	Aspirations in 8th grade: at most vocational school				
	(1)	(2)	(3)	(4)	(5)
Repeated	0.387***	0.278***	0.169***	0.0702***	0.0732***
	(0.011)	(0.011)	(0.0097)	(0.0096)	(0.0094)
Maximum years of education of		-0.0358***	-0.0195***	-0.0141***	-0.0111***
parents		(0.00056)	(0.00038)	(0.00029)	(0.00028)
Repeated × Maximum		-0.0376***	-0.0299***	-0.0279***	-0.0275***
years of education of parents		(0.0037)	(0.0035)	(0.0034)	(0.0034)
6th grade aspirations, baseline: at most vocational					
High school diploma or			-0.384***	-0.296***	-0.280***
post-secondary non-tertiary			(0.0031)	(0.0030)	(0.0029)
College or university			-0.437***	-0.301***	-0.279***
			(0.0033)	(0.0031)	(0.0030)
Constant	0.0991***	0.100***	0.453***	0.822***	0.772***
	(0.0019)	(0.0012)	(0.0033)	(0.032)	(0.032)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Parental education is demeaned. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

C.0.3.2 Using continuous SES

10.14754/CEU.2022.08

	Aspirations in 8th grade: tertiary				
	(1)	(2)	(3)	(4)	(5)
Repeated	-0.428***	-0.340***	-0.154***	-0.0494***	-0.0454***
	(0.0072)	(0.010)	(0.0094)	(0.0088)	(0.0087)
Maximum years of education of		0.0948***	0.0514***	0.0398***	0.0331***
parents		(0.00032)	(0.00034)	(0.00033)	(0.00031)
Repeated × Maximum		-0.0371***	-0.0244***	-0.0143***	-0.0126***
years of education of parents		(0.0039)	(0.0033)	(0.0031)	(0.0030)
6th grade aspirations, baseline: at most vocational					
High school diploma or			0.148***	0.0537***	0.0451***
post-secondary non-tertiary			(0.0021)	(0.0019)	(0.0019)
College or university			0.630***	0.399***	0.377***
			(0.0024)	(0.0026)	(0.0026)
Constant	0.541***	0.537***	0.164***	-0.373***	-0.313***
	(0.0043)	(0.0018)	(0.0022)	(0.025)	(0.026)
Observations	444632	444632	444632	444632	444632
6th grade controls	no	no	no	yes	yes
Year fixed effect	no	no	no	yes	yes
School fixed effect	no	no	no	no	yes

Standard errors are clustered on school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Probability of 8^{th} grade aspirations being at most vocational education

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Parental education is demeaned. In column 3, 6th grade aspirations are added as controls. In column 4, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 5 includes school fixed effects.

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated	-2.462***	-1.904***	-0.983***	-1.343***	-0.528***	-0.531***
	(0.043)	(0.052)	(0.045)	(0.057)	(0.054)	(0.053)
Maximum years of		0.415***	0.194***	0.196***	0.140***	0.115***
education of parents		(0.0022)	(0.0017)	(0.0017)	(0.0015)	(0.0014)
Repeated ×		-0.0509***	0.00187	0.132***	0.112***	0.116***
Maximum years of education of parents		(0.019)	(0.017)	(0.022)	(0.021)	(0.021)
6th grade			0.528***	0.525***	0.341***	0.319***
aspirations in years			(0.0019)	(0.0019)	(0.0019)	(0.0019)
Repeated ×				-0.162***	-0.0862***	-0.0844***
6th grade aspirations in years				(0.021)	(0.020)	(0.019)
Maximum years of				-0.00544***	-0.00379***	-0.00437***
education of parents × 6th grade aspirations in years				(0.00069)	(0.00063)	(0.00062)
Repeated ×				0.0468***	0.0294***	0.0301***
Maximum years of education of parents × 6th grade aspirations in years				(0.0071)	(0.0067)	(0.0067)
Constant	14.71***	14.70***	14.65***	14.66***	11.57***	11.69***
	(0.019)	(0.0081)	(0.0046)	(0.0056)	(0.15)	(0.15)
Observations	391412	391412	391412	391412	391412	391412
6th grade controls	no	no	no	no	yes	yes
Year fixed effect	no	no	no	no	yes	yes
School fixed effect	no	no	no	no	no	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: Educational aspirations in 8^{th} grade in years

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. SES is proxied with the years of education of the highest educated parent. Both the 6th grade aspirations variable and the parental education variable are demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

	8th grade educational aspirations in years					
	(1)	(2)	(3)	(4)	(5)	(6)
Repeated 7th grade	-2.826***	-1.938***	-1.034***	-1.751***	-0.790***	-0.801***
	(0.078)	(0.071)	(0.068)	(0.16)	(0.15)	(0.15)
High SES		2.255***	1.054***	1.092***	0.643***	0.501***
		(0.023)	(0.015)	(0.017)	(0.014)	(0.014)
Repeated 7th grade		-0.327*	-0.00860	0.680***	0.673***	0.617**
× High SES		(0.19)	(0.16)	(0.25)	(0.25)	(0.24)
6th grade			0.536***	0.500***	0.331***	0.316***
aspirations in years			(0.0027)	(0.0047)	(0.0046)	(0.0046)
Repeated 7th grade				-0.210***	-0.123***	-0.121***
× 6th grade aspirations in years				(0.044)	(0.042)	(0.042)
High SES ×				0.0543***	0.0718***	0.0644***
6th grade aspirations in years				(0.0054)	(0.0051)	(0.0050)
Repeated 7th grade				0.163*	0.0690	0.0682
× High SES × 6th grade aspirations in years				(0.085)	(0.079)	(0.078)
Constant	15.27***	13.80***	14.57***	14.52***	10.61***	10.90***
	(0.024)	(0.017)	(0.012)	(0.015)	(0.29)	(0.29)
Observations	194852	194852	194852	194852	194852	194852
6th grade controls	no	no	no	no	yes	yes
Year fixed effects	no	no	no	no	yes	yes
School fixed effects	no	no	no	no	no	yes

Standard errors are clustered on the class level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: Educational aspirations in 8^{th} grade in years, Admin3 data

Note: The sample includes children who were in primary school in 6th grade and have not repeated the 6th grade. Standard errors are clustered on 6th grade school level. The 6th grade aspirations variable is demeaned. In column 4, 6th grade aspirations are added as controls. In column 5, 6th grade controls are: gender, whether the student has special education needs, year in which the student took the 6th grade test, 6th grade mathematics test score, 6th grade reading test score, mother's labor market status, father's labor market status, 6th grade midterm mathematics, literature, Hungarian grammar, effort, and behavior grade. For all variables, dummies are used for missing values. Column 6 adds 6th grade school fixed effects.

C.0.3.3 Regression of continuous aspirations on the Admin3 data

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.180	-0.139	-0.239
	(0.31)	(0.31)	(0.34)
High SES	0.849***	0.631***	0.572***
	(0.15)	(0.15)	(0.16)
Repeated ×	-0.258	-0.313	-0.0608
High SES	(0.59)	(0.57)	(0.64)
6th grade		0.280***	0.244***
aspirations in years		(0.048)	(0.055)
Repeated ×			-0.122
6th grade aspirations in years			(0.17)
High SES ×			0.124
6th grade aspirations in years			(0.091)
Repeated ×			-0.0762
High SES × 6th grade aspirations in years			(0.37)
Observations	2772	2772	2772
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.11: 8^{th} grade aspirations for those failing mathematics at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

C.0.3.4 Regressions on subsamples with health controls

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.512***	-0.459***	-0.547***
	(0.083)	(0.079)	(0.088)
High SES	0.746***	0.555***	0.540***
	(0.027)	(0.026)	(0.026)
Repeated ×	0.354	0.405*	0.522**
High SES	(0.22)	(0.21)	(0.21)
6th grade		0.260***	0.236***
aspirations in years		(0.0083)	(0.010)
Repeated ×			-0.0818**
6th grade aspirations in years			(0.036)
High SES ×			0.0601***
6th grade aspirations in years			(0.015)
Repeated ×			0.0473
High SES × 6th grade aspirations in years			(0.13)
Observations	28193	28193	28193
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.12: 8^{th} grade aspirations for those just passing mathematics at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	8th grade aspirations in years		
	(1)	(2)	(3)
Repeated	-0.522***	-0.309*	-0.566
	(0.17)	(0.16)	(0.36)
High SES	0.769***	0.476***	0.523***
	(0.015)	(0.014)	(0.016)
Repeated ×	0.0109	-0.143	0.165
High SES	(0.33)	(0.30)	(0.46)
6th grade		0.378***	0.337***
aspirations in years		(0.0030)	(0.0054)
Repeated ×			-0.0605
6th grade aspirations in years			(0.089)
High SES ×			0.0581***
6th grade aspirations in years			(0.0059)
Repeated ×			0.0883
High SES × 6th grade aspirations in years			(0.13)
Observations	157406	157406	157406
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.13: 8^{th} grade aspirations for those with higher mathematics marks at 6^{th} grade midterm, Admin3 database

Note: Controls include gender, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0986	-0.0799	-0.189*
	(0.13)	(0.12)	(0.11)
High SES	0.220***	0.195***	0.204***
	(0.064)	(0.064)	(0.065)
Repeated ×	-0.169	-0.228	-0.202
High SES	(0.28)	(0.29)	(0.31)
6th grade		0.0415***	0.0515***
aspirations in years		(0.015)	(0.019)
Repeated ×			-0.143***
6th grade aspirations in years			(0.050)
High SES ×			-0.0216
6th grade aspirations in years			(0.029)
Repeated ×			0.238
High SES × 6th grade aspirations in years			(0.18)
Observations	1434	1434	1434
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.

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

Table C.14: Attending a secondary school ending with a high school diploma (1 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

10.14754/CEU.2022.08

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.0761*	-0.0674*	-0.0717
	(0.042)	(0.041)	(0.052)
High SES	0.158***	0.132***	0.134***
	(0.0084)	(0.0084)	(0.0085)
Repeated ×	-0.0139	-0.0179	-0.0195
High SES	(0.070)	(0.068)	(0.075)
6th grade		0.0385***	0.0420***
aspirations in years		(0.0020)	(0.0029)
Repeated ×			-0.00619
6th grade aspirations in years			(0.026)
High SES ×			-0.00725**
6th grade aspirations in years			(0.0037)
Repeated ×			0.0148
High SES × 6th grade aspirations in years			(0.030)
Observations	17540	17540	17540
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.15: Attending a secondary school ending with a high school diploma (2 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

	Attending sec. school giving high school diploma		
	(1)	(2)	(3)
Repeated	-0.224***	-0.220***	-0.107
	(0.073)	(0.072)	(0.072)
High SES	0.0745***	0.0678***	0.0492***
	(0.0023)	(0.0023)	(0.0021)
Repeated ×	0.116	0.110	-0.00263
High SES	(0.093)	(0.092)	(0.090)
6th grade		0.00865***	0.0264***
aspirations in years		(0.00030)	(0.00074)
Repeated ×			0.0279**
6th grade aspirations in years			(0.014)
High SES ×			-0.0247***
6th grade aspirations in years			(0.00078)
Repeated ×			-0.00121
High SES × 6th grade aspirations in years			(0.021)
Observations	114298	114298	114298
Controls	yes	yes	yes
Year fixed effect	yes	yes	yes
School fixed effect	yes	yes	yes

Standard errors are clustered on the school level. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.16: Attending a secondary school ending with a high school diploma (3-5 in mathematics, Admin3 database)

Note: Controls include gender, whether the student has special education needs, 6th year mathematics and reading test scores, midterm literature, grammar, effort, and behavior grades, whether the parents were together in 6th grade, whether they separated between 6th and 8th grades, number of visits to the general practitioner in the schoolyear and days spent in a hospital in the schoolyear. For all variables, dummies are used for missing values. 6th grade aspirations are demeaned.

C.0.4 Data Appendix

The student level questionnaires and test score data are available between 2008 and 2017, with in total around 2.8 million observations. I am interested in those students for whom at least 6^{th} and 8^{th} grade data can be linked, so as a first step, I drop all students whose data cannot be matched between these two grades. Then, I create an indicator variable for whether the student repeated any grade up to the 6^{th} grade. I use two variables here: whether the student repeated in the lower grades of primary school, and whether she repeated in the higher grades of primary school, both reported in grade 6. I classify the student as no repeater if both variables equal 1, which means no repetition in those particular grades. Unfortunately, in 6^{th} grade, 20 percent of this constructed measure are missing. I recover this data from the 8^{th} grade and 10^{th} grade surveys, wherever possible, since later surveys also include these questions. When reporting about repetition in lower grades was inconsistent through survey years, I use the mode of the answers to impute the 6^{th} grade value. In the end, I have self-reported information on repetition up to $6th$ grade from 94.3 percent of the 6^{th} grade sample. Because of the compulsory school starting age rules in Hungary, students are 12 or 13 years old in the year they finish 6^{th} grade (that is the time of the test), but in a small percentage of cases it is possible that students – without repeating a grade – turn 14 in this year. Therefore, I drop students younger than 12 when writing the 6^{th} grade test, and assign students older than than 14 to the repeater group.[30] This is around 1 percent of the sample. I construct a wide database where the time variable is the grade, and I keep the first observations from 6^{th}, 8^{th} and 10^{th} grades. I define a variable for each grade indicating if the student had multiple observations from that particular grade, meaning that she repeated it, and another variable that counts how many observations the student has from the grade.

C.0.4.1 Birth year and month and school starting age variables The source of birth year and month can be regarded as administrative data, but it contains missing values across years (less than 1 percent of all observations). I recover missing data from the other years' data of the same person. School starting age is a constructed variable from the self-reported year when students started school and their birth year and month. School starting age means the age the child had already turned when started school. Datasets before 2015 only contain this constructed school starting age, while from 2015, they also include the school starting year as students reported it. For both variables, I first create a corrected variable that adds the last non-missing value to all observations of the same student. Then, as later I would only like to use the school starting year, I calculate school starting year from the corrected school starting age variable for students for whom the year was missing. As school starts in September, I regard students up to September as having turned the age they turn in that year when they started school, while students born between October and December, their school starting age is a year younger than the age they turn in the particular year.

[30] I use the first occurrence in the 6^{th} grade sample, since a fraction of students had to repeat 6^{th} grade, which resulted in them appearing multiple times in the 6^{th} grade samples.

C.0.4.2 Creating years of education from the categorical education variables I create continuous variables from two categorical education variables: parental education and educational aspirations. For both variables I use the following coding: unfinished primary education = 7 years, primary education = 8 years, vocational and technical secondary education = 11 years, high school diploma = 12 years, upper-secondary non-tertiary qualification = 14 years, college or BA degree = 15 years, university or MA degree and doctoral degree = 17 years. I chose to code a doctoral degree similarly as a master's degree because in primary school children probably cannot apprehend the length of a doctoral education, so setting their aspirations to 20-22 years of schooling would be a large overestimation of their educational preferences.

C.0.4.3 Admin3 NABC extension I took the same data cleaning steps as for the National ABC until keeping the last observations and the reshape wide. I use all observations for the admin data and keep the data in a long format to be able to merge + append with the admin panel. I first matched the observations from the National ABC extension to data from May in the same year in the administrative database. I dropped about 280 observations where the matches were wrong: these ID-s belonged to much older people in the Admin than in the National ABC. I dropped everyone who were older than 22 (the highest age in the National ABC) in the matched database according to the administrative database. I checked if the rest of the observations are matched well in terms of age: I considered a match bad if KOR from the Admin3 database were higher than AGEATTEST + 1 or lower than AGEATTEST - 1 from the National ABC database. First I corrected the birth year for those observations where the age mismatch was the result of inconsistent reporting of the birth year. At the end I still had 40 observations with age mismatch. I dropped these 40 observations, too. I did the same steps as with the National ABC database, but here some data were missing: there was no data on class and on special education needs and valid/ not valid test status. I merged the National ABC data to the monthly main Admin3 database by setting the NABC month as May. Then I merged the monthly healthcare data to the main file. For the estimations, I created school-year aggregates of the monthly healthcare data, e.g., number of visits to the GP in the 6^{th} grade, 7^{th} grade, etc. I kept the first observations of 6^{th} grade, 8^{th}, and 10^{th} grade NABC data and the corresponding data from Admin3 from these years. I also kept the 7^{th} grade healthcare data for predicting retention.