Essays in Experimental and Education Economics

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

vorgelegt von Mark Jeffrey Spils im Frühjahrs-/Sommersemester 2023.

Abteilungssprecher:	Prof. Klaus Adam, Ph.D.
Referent:	Prof. Dr. Henrik Orzen
Korreferent:	Prof. Dr. Katja Maria Kaufmann

Tag der mündlichen Prüfung:	28. März 2023

Acknowledgements

First and foremost, I would like to thank my advisors Henrik Orzen and Katja Kaufmann for their continuous support throughout the entirety of my Ph.D. studies. I am grateful for their valuable advice, their comments and feedback on my work and their continuous availability. Our meetings were incredibly interesting and helped to develop both my research and myself as a researcher. Moreover, I thank Katja for her direct contributions to Chapter 1 as co-author.

I am also thankful for discussions with fellow researchers at the University of Mannheim and other institutions, especially the University of Bonn, throughout my Ph.D. studies. Their comments and suggestions helped with improving the chapters in this dissertation. I want to give special thanks to Xi Lin and Jonathan Stäbler for the countless inspiring conversations at the time when we were sharing offices and to the lovely people from the experimental reading group for all the interesting and fun discussions and lunch meetings.

For Chapter 1, I gratefully acknowledge funding for the administrative microdata by the German Research Foundation (DFG) through CRC TR 224 (Project C01). For Chapters 2 and 3, I am grateful for the funding for my experiments by the chair of Experimental Economics. Moreover, I am most appreciative of the efforts from (former) mLab managers Andrej Svorenčik and Franziska Heinicke by facilitating experimental sessions at times when it was most difficult. More generally, I thank the Graduate School of Economic and Social Sciences at the University of Mannheim for providing financial support and the research oriented environment.

Ultimately, I would like to express my deepest gratitude to my parents, Bart and my friends (in particular Tim and Ronald). They have been a continuous support and my sounding board during my 6.5 years in Mannheim. They helped me to persevere during the most difficult times. Lotte, thank you for all your love, patience and support. All this would not have been possible without you.

Contents

List of Tables

List of Figures

[This page intentionally left blank]

Preface

This dissertation consists of three chapters which are self-contained. However, there is a common denominator between all three chapters. Each chapter looks at unintended indirect effects of policy interventions. Chapter 1 is a continuation of my dissertation proposal (Spils, 2018) and explores the long-run effects of reducing the number of STEM-hours in high school. Chapter 2 studies the effect of a buyer's bias for an insurance product on the market equilibrium. Chapter 3 investigates the economic effects of wearing a face mask in the laboratory during various well-known tasks from the literature.

Chapter 1 makes use of administrative microdata from Statistics Netherlands (CBS), which allows us to investigate a policy change in the Netherlands which makes it easier for students to choose the STEM field in secondary school. As the reduction in study load is much larger in one secondary school track than in the other, this policy change is considered a quasi-experiment. Chapter 2 analyzes an online experiment where buyers and sellers are interacting on a market with two insurance products. The experimental treatment is a difference in when both products are presented to the buyers. This experiment can be used to see both to what extent buyers are affected by the different treatments and if the sellers respond to different preferences of the buyers. Chapter 3 uses a laboratory experiment where subjects do ten tasks either with or without face mask, according to the local laws and university guidelines regarding the COVID-19 pandemic. Hereafter, a brief summary of each chapter is provided.

In Chapter 1, which is joint work with Katja Kaufmann, we analyze the short- and long-run effects of a reduction in the mandatory STEM hours of the STEM field/specialization in Dutch secondary schools. In particular, the work load for the STEM field, which is a prerequisite for enrolling in STEM majors at university/technical college, decreased more strongly in the academic track than in the technical track. Employing a difference-in-difference approach combined with Dutch administrative data, we find that the policy led to a significant increase in the take-up of the STEM field in high school, especially for women and students from higher income households. In the longer-run, however, enrollment in STEM majors at university did not increase. In fact, after the policy change previously underrepresented groups such as women and individuals from low-income families were even relatively less likely to pursue a STEM degree. The decrease of women graduating from STEM was primarily driven by women with STEM parent(s), suggesting that it was due to negative signals about their preparedness for a STEM major.

In Chapter 2, I investigate a market for mandatory insurance with a choice between a low deductible (product Low) and a high deductible (product High). I describe theoretically how a bias for a product affects the market equilibrium. If buyers have an initial (status-quo) bias for product Low, then this leads to an overvaluation for product Low compared to standard economic preferences. As a consequence, some buyers purchase different products compared to standard decision making. If the bias is leads to an homogeneous additional utility value to the valuation of all buyers, both products become cheaper as only the marginal buyer is switching to product Low. If the bias affects only a subset of the buyers at random and if biased buyers always buy product Low, also buyers with a low risk of getting a damage buy product Low. This lowers the price of product Low more, but does not lower the price of product High. This lowers the price difference and would make product Low even more attractive. I conducted an online experiment to see if the equilibrium over time is affected by a bias for either product. I find that there was a bias for product Low, but not for product High. Moreover, I find that even though the bias for product Low weakened over time, the sellers' beliefs about the market composition immediately changed which can lead to a distorted equilibrium in the long run.

In Chapter 3, I analyze the effect of wearing disposable face masks on economic behavior using a laboratory experiment. I investigate the effects of wearing a face mask for both a short period of time and wearing it continuously for 45 minutes longer. Subjects in half of the sessions had to wear a face mask according to the local law. In the other half of the sessions, subjects did not have to wear a mask during the experiment. There are no significant differences in cognition, risk aversion, loss aversion or social preferences between subjects with and without face masks, with the exception of higher offers in the ultimatum game. There is a significant reduction of productivity in a short high-paced task, but not in a longer task where participants could pace their efforts.

1 The Long-Run Effects of a Reduction of STEM Hours in High School

with Katja Kaufmann

1.1 Introduction

Technological progress in recent decades strongly suggests that future economic growth can primarily be expected in the fields of science, technology, engineering and mathematics (STEM) (OECD, 2010). In Europe, the demand for STEM graduates is expected to grow by 9.1 percent between 2020 and 2030 (Cedefop, 2020) and in the U.S. even by 10.8 percent between 2021 and 2031 (BLS, 2021) (compared to 4.4 and 4.9 percent for all other occupations, respectively). It is a worldwide challenge to keep up with this growing demand. According to the U.S. Department of Education (2017), only 18 percent of Bachelor degrees were awarded in STEM fields in 2016. This is partly due to an underrepresentation of women in STEM fields. Despite the fact that women received 58 percent of all Bachelor degrees, they only received 36 percent of all STEM Bachelor degrees. Besides women, also minorities and students from less privileged households are underrepresented in STEM fields worldwide (Griffith (2010); Kokkelenberg and Sinha (2010)). Students from these groups are less likely to choose a STEM major in university and the ones that do are more likely to drop out (Chen and Soldner, 2014). One particular focus among policy makers is therefore to enact policies aimed at tapping the unused potential especially among the underrepresented groups in order to increase the supply of STEM graduates.

Why is the fraction of students graduating in STEM fields so low, despite the fact that STEM graduates have higher earnings than graduates from other fields (e.g. Abramitzky et al. (2019), Altonji et al. (2016), Ardicia-cono (2004), James et al. (1989)) and despite the fact that vacancies in this sector are many and foreseen to increase further? Beside monetary reasons to opt for a STEM major, another potential explanation is linked to preferences. Students face a utility maximization problem when deciding about their college major and both monetary and non-monetary benefits and costs are taken into account. For example, Zafar (2013) finds that women have stronger preferences for non-STEM fields. Related to this point, students might opt for non-STEM majors because of higher effort costs of obtaining a STEM major which might outweigh the long run benefits of better prospects in the labor market. While studies suggest that women have lower effort costs of studying, another reason for opting out of STEM might be that women care more about achieving good grades and thus opt out of STEM fields where grade averages are lower (Ahn et al., 2019).

In this paper, we analyze the short- and long-run effects of a curriculum change in Dutch secondary schools, more specifically a reduction in the mandatory STEM hours of the STEM field/specialization. Graduating from high school with STEM specialization is a prerequisite for enrolling in a STEM major. This holds for students graduating from high school in the academic track (VWO), which is the requirement for enrolling in a research university, as well as for students graduating from the technical track in high school (HAVO) necessary for enrolling in a university of applied sciences. To increase the accessibility and attractiveness of the STEM field (*Nature/Tech*) in secondary school, the Dutch government lowered the work load in terms of field-specific course hours starting in 2007, and the reduction was particularly strong in the academic track. This led to more students choosing the STEM field *Nature/Tech*, in particular in the academic track, and consequently more students meeting the prerequisites to enroll in a STEM major at university.

More specifically, the number of mandatory hours in field-specific courses (such as math and physics in *Nature/Tech*) were lowered by 17.5 percent in the academic track (compared to 6.9 percent in the applied university track), while the number of hours of freely elective subjects increased to fully compensate for the drop in hours in field-specific courses. Since study load and effort costs tend to be highest for quantitative subjects, such as mathematics and physics, we expect the decrease in field-specific hours to have the largest effects on study load and effort costs in the STEM field *Nature/Tech*. With the exception of the differential drop in field-specific hours, the two tracks (academic/research university track and technical/applied university track) resemble each other in important ways, such as in terms of students having to choose a field/specialization to graduate in and only being able to enroll in a STEM major at university if students graduated from high school with the field *Nature/Tech*. Moreover, the fraction of students choosing *Nature/Tech* developed in the same way in both tracks in the years prior to the reform, consistent with the parallel trend assumption underlying the difference-in-differences (DID) approach. Therefore, we apply the DID methodology to analyze the effect of the reduction in mandatory STEM hours in the STEM field in the short-run (i.e. graduating from high school with *Nature/Tech* and thus satisfying the formal requirement for enrolling in a STEM major at university) and also in the longer-run. In particular, we investigate the effect on the probability of graduating with a STEM bachelor degree and STEM master degree. Moreover, we analyze the policy's impact on other long-run outcomes, such as labor market earnings and marriage and fertility outcomes when students are in their late twenties/early thirties.

To analyse the causal effect of lowering the effort costs for graduating high school with the *Nature/Tech* field, we use Dutch administrative microdata from Statistics Netherlands (CBS). This dataset contains information on the entire Dutch population in terms of their family background, education histories (including the track in secondary school, field choice, academic grades/GPA in secondary school, college major choice as well as highest obtained degree in college), labor market outcomes and marriage and fertility outcomes. The first cohort after the policy change is currently observed until age 29. Thus we can follow students for nearly 15 years after their choice of a field/specialization and for more than ten years after graduating from secondary school, allowing us to observe their full educational histories including bachelor and master degrees, and labor and family formation outcomes until students' late twenties/early thirties.

In our analysis we first investigate the short-run effect of the reduction in mandatory STEM hours on the likelihood of graduating from high school with *Nature/Tech*. The policy increased the likelihood of specializing in *Nature/Tech* by 11 percentage points (from a baseline of 17 percent in the academic track) and thus substantially increased the fraction of students satisfying the formal requirements to enrol in a STEM major at university.

Since there is a particular interest among policy-makers to increase the number of students from underrepresented groups, we also analyze the short-run effects on subgroups of the population. The direct effect of the policy change is stronger for female than for male students. Women's likelihood of graduating high school with *Nature/Tech* increased by 14 percentage points compared to only 7 percentage points for men. The policy thereby reduced the gender gap by nearly 7 percentage points (from 23 percentage points to 16 percentage points for students in the academic track) and consequently the prospective gender gap in terms of STEM enrollment at university. In terms of socioeconomic background on the other hand, the policy increased the gap between low- and high-income students. While students from less privileged households increased the likelihood of graduating with *Nature/Tech* by 7 percentage points, students from more privileged backgrounds increased the likelihood significantly more (by 11 percentage points), thereby increasing the socioeconomic status gap. Thus, in the short-run, the policy (partially) met the intended goals: Overall a substantially larger fraction of students met the formal requirement to enrol in a STEM field at university and the gender gap decreased. On the other hand, the socioeconomic status gap increased somewhat.[1]

However, the effects we find in the medium and longer-run paint a different picture. Despite the fact that the fraction of students satisfying the formal requirements for STEM at university went up substantially, the likelihood of enrolling into or graduating with a STEM bachelor or master remains unchanged. While the policy led to a slight increase in terms of male students graduating with a STEM degree (by 1.4 percentage points in terms of STEM bachelor and by 1 percentage point in terms of STEM master), the effect on women is significantly smaller and, more specifically, there is no increase (or even a slight decrease) for women graduating with a STEM degree. Thus in the longer-run the policy led to a widening of the existing gender gap in STEM graduates, contrary to what was intended. Also the socioeconomic status gap increased in response to the policy. Already in the short-run low-income students increased their take-up of *Nature/Tech* in high school to a significantly lower extent than high-income students. In the longer-run in terms of graduating with a STEM major the gap not only increased, but low-income students were significantly less likely to obtain a STEM master degree in response to the policy.

What are the underlying mechanisms behind the observed short-run and long-run effects of the policy? In the short-run, the reduction in mandatory STEM hours (in particular in math and physics) in the STEM field *Nature/Tech* implied an important decrease in the effort costs of graduating from high school in *Nature/Tech*. The policy most likely also raised the expected GPA of graduating in *Nature/Tech*, since grades now depended to a smaller extent on the performance in STEM subjects, which tend to be graded more strictly such that average grades in these subjects tend to be worse. At the same time obtaining a high school degree with STEM field has the following advantages: the option value of satisfying the formal requirements to choose any college major including STEM, improved skills in quantitative subjects and a more prestigious high school degree which can have direct benefits in the labor and marriage market (even without university attendance or STEM at university), and an increased opportunity to have a social network or find a spouse/partner with a high school

[1] We also conducted a heterogeneity analysis by migration background, using different classifications, but we do not find any significant differences in short- or long-run effects by migration background.

degree in STEM (and thus have friends/spouse/partner with higher labor market potential and career opportunities).

Since the curriculum change had the direct implication of lowering the costs of graduating with *Nature/Tech*, we would expect students to increase the take-up of *Nature/Tech* in high school. It is less clear what to expect in the longer-run and what to expect for groups previously underrepresented in STEM, such as women and low-income students. The idea underlying the policy was to draw students into STEM and give them the opportunity to experience and gain interest in the STEM field, in particular students who had less exposure to STEM and fewer role models prior to the reform, so that the hope was to reduce gaps in terms of gender and socio-economic status. In the short-run this turned out to work, at least for women who increased their take-up of STEM in high school more strongly than men. In the longer-run however, while there was a slight increase in male high school graduates and high school graduates from higher-income families who took up the opportunity to pursue a STEM degree in college, there was no change or even a decrease in terms of female students and lower-income students graduating from college with a STEM major. Why did women and low-income students not increase or even decrease their likelihood to graduate with a STEM degree?

One important direct implication of the policy is that high school students graduating in *Nature/Tech* have acquired less field-specific knowledge, in particular in terms of math and physics, than prior to the reform. It is well-known that women and men differ in terms of self-confidence and that men tend to be more self-confident or even overoptimistic with respect to their abilities (i.e.Morin (2015), Preckel et al. (2008), Niederle and Versterlund (2007)). Thus, one possible explanation for women to decrease their likelihood of pursuing STEM in college relative to men is related to the fact that STEM high school graduates are certainly less well prepared for pursuing a STEM degree in college. Worse preparation implies a higher likelihood of drop-out and worse expected grades, which women seem to be particularly concerned about, while being less confident in their own abilities. Similarly, students from less privileged backgrounds tend to be less self-confident (Guyon and Huillery, 2021). Moreover, they are less able to compensate the reduction in STEM hours/preparation, for example via private tutoring or help of the parents, since few low-income students have parents with a college degree, let alone with a STEM degree.

To investigate the role of having parents who have the resources or skills to compensate for the lack in preparedness (via remedial tutoring or direct help), we analyze heterogeneous effects in terms of policy impact on male and female students by whether at least one of their parents have a college degree or a STEM degree. On the one hand, students with more highly educated parents might be better able to compensate and learn the necessary skills outside of high school or university. On the other hand, more educated parents and in particular parents with STEM degrees, are also more aware that their children lack important abilities and knowledge for pursuing STEM, which renders a STEM major in college even more costly than prior to the reform.

Our results suggest that the decision of women not to pursue a STEM degree in college in response to the reform was linked to the educational background of their parents. We find that the curriculum change had particularly negative effects on women with STEM parent(s), that is women with at least one STEM parent are 3.8 percentage points less likely to obtain a STEM bachelor and 2.4 percentage points less likely to obtain a STEM master in response to the policy change compared to women without STEM parents. The same does not hold for women with college-educated parents. This suggests that women with STEM parents receive the signal from their parents that they lack fundamental abilities/knowledge for pursuing a STEM degree in college. This makes the pursuit of a STEM degree not only more costly in terms of study effort than prior to the reform, but it importantly also leads to the expectation of worse grades and a higher risk of drop-out. While for male students with STEM parents, the returns to STEM appear to still outweigh the increase in study costs (possibly also due to higher self-confidence/overconfidence and thus higher expected grades and/or a lesser concern with worse grades, see Ahn et al. (2019)), the same is not true for female students (whose returns to a STEM degree might also be lower due to a higher likelihood of working part-time and of having career interruptions due to children).

These results indicate that lowering the prerequisites in high school to enroll in a STEM major will ultimately not lead to more female STEM graduates. Female students need stronger signals of mathematical ability to choose male-dominated STEM subjects, even when they have the same grades (Justman and Méndez, 2018). Therefore, while lowering STEM prerequisites in high school appears to induce women to increase the take-up of the STEM field in high school, ultimately such a policy backfires and reduces the number of female STEM graduates at university, because women feel less prepared in the relevant STEM subjects.

Why did women increase the take-up of *Nature/Tech* in high school in the first place, and more strongly than men? First, we show that women (and high-income students) responded more strongly to the policy, the higher the fraction of high-income students in *Nature/Tech* already prior to the reform, i.e. the effect on women and high-income students appears to have been reinforced via peer effects.[2] This might be one factor for why we see a stronger increase in the take-up of *Nature/Tech* in high school among women and high-income students. How did the change in social network and in the pool of potential partners affect women's long-run outcomes? What are the effects of the policy on long-run labor market and family formation outcomes for women with STEM parents, who reduced their likelihood of graduating from college with STEM bachelor or master in response to the policy? We find that the policy led to an increase in labor earnings of women with STEM parents, potentially driven by better quantitative skills (due to *Nature/Tech* in high school, albeit a reduced likelihood of STEM bachelor or master) or by a stonger social network in high school. Moreover, the likelihood of having a spouse also increased (and the likelihood of children increased relative to women without STEM parents, but not in absolute terms). While it is certainly insightful to investigate even longer-run effects on labor market and marriage outcomes of students into their mid-/late-thirties, by their late twenties most students have already been in the labor force for several years (even if they have a bachelor or master degree) and a large fraction of people is married. Thus our results suggest that the long-term effect of choosing *Nature/Tech* in high school, but opting out of a STEM bachelor or master, appears to have had positive or at least no negative long-run effects on women with STEM parents, consistent with them anticipating the lack of preparedness after the curriculum change.

Our paper is related to the following strands of the literature: Closest to our paper is a recent small, but growing, literature investigating the effects of curriculum changes in STEM subjects in high school on the STEM major choice. Biewen and Schwerter (2019) make use of a policy change in Germany making math compulsory in the last two years of high school, which increases the share of STEM students and increases the gender gap. De Philippis (2021) analyzes the effect of a policy change in the UK that led to more schools offering advanced science in high school and finds an increase in the share of STEM students and a widening of the gender gap.

Compared to these papers we analyze what are the short- and long-run

[2]Interestingly, we do not find that female students increased the take-up of *Nature/Tech* more strongly, the larger the fraction of female students prior to the reform.

effects on STEM and other outcomes of **reducing** the number of mandatory STEM hours. Investigating the impact of this policy change is interesting for at least three main reasons. First, effects might not be monotonous, thus the impact of a reduction in mandatory STEM hours might not be equivalent to the negative of the effect of an increase. Second, the two existing papers look at making math hours mandatory or at the effect of more schools offering/introducing advanced science classes, so the marginal individuals affected by these policies are likely to be quite different. Third, in the light of the fact that both analyzed policies led to increases in the gender gap, it is interesting and highly policy relevant to understand the effect of a policy which had the goal to increase STEM exposure and draw students into STEM, in particular among previously underrepresented groups such as women and students from less privileged households. Moreover, we provide evidence not only in terms of short- and long-run STEM outcomes, but also on labor market and family formation outcomes, and we show results for further subgroups, such as for students from different socio-economic and migration backgrounds. Lastly, we are able to shed some light on the underlying mechanisms, by investigating - among other aspects- effects depending on whether or not students' parents have a STEM or college degree and how effects vary depending on the peer group.

Another related literature investigates the effect of changes in math curriculum. Joensen and Nielsen (2009) and Joensen and Nielsen (2016) analyze the effect of a curriculum change in Denmark, which allows students to combine advance math with biology for graduating with a STEM field. They find that the policy increased education and earnings and led women to take more intensive math subjects and more competitive careers decreasing gender gaps. We complement their findings by evaluating a different type of policy, providing more direct evidence on long-run STEM outcomes and by analyzing the heterogeneity of effects by socioeconomic status and migration background. Two other papers evaluating the effects of changes in math instruction time on disadvantaged groups (low-skilled 9th graders and African Americans) find positive effects on educational outcomes and earnings (see Cortes et al. (2015) and Goodman (2019)).[3] Lavy (2015) and Abramitzky et al. (2019) analyze the effects of instruction time of different school subjects on educational achievement.

Also related to our paper is the literature on college major choice (see,

[3]Two papers that are not evaluating policy changes are Delaney and Devereux (2019) and Aucejo and James (2021) who investigate the relevance of math and verbal skills for university enrollment and performance and in the former case also on the likelihood of acquiring a STEM degree.

among others, Zafar (2013), Altonji et al. (2016)) and more specifically on the choice of a STEM major (see, e.g., Ahn et al. (2019)). Our paper contributes to this literature by showing the relevance of curriculum changes on STEM major choice overall and for different subgroups.

Lastly, our paper is linked to a literature analyzing different types of policies aimed at decreasing the gender gap in STEM, for example by providing students with female role models (see, e.g., Bettinger and Long (2005), Carrell et al. (2010) and Breda et al. (2020)). We show that the policy of making access to the STEM field easier by decreasing mandatory STEM hours raises take-up and decreases the gender gap in the short-run, but backfires in the longer-run by actually increasing the gender gap in terms of graduating with a STEM bachelor or master.

This paper is organized as follows. Section 1.2 discusses the institutional framework in the Netherlands and describes the policy change. The following sections 1.3 and 1.4 describe the empirical model, the data used in the analysis and provides descriptive statistics. Section 1.5 analyzes the results and section 1.6 explores the mechanisms. Concluding remarks are offered in section 1.7.

1.2 Institutional framework

In this section, we describe the system of secondary education in the Netherlands, the enrollment in tertiary education and the policy change and how it affected students in the different tracks of secondary school.

1.2.1 Secondary education in the Netherlands

At age 12, students finish primary school and go to (mandatory) secondary school. The student enrolls into one of three tracks of secondary education: the six-year pre-university education (VWO), five-year higher secondary vocational education (HAVO) or four-year intermediate secondary vocational education (VMBO). The allocation of students to these tracks is based on the primary school teachers' recommendation and centralized test-scores. The secondary school itself can be chosen freely by the student and the parents conditional on the school offering the student's track. From now on, we will only look at HAVO and VWO as only these tracks were affected by the policy. HAVO and VWO are also the only tracks which give access to higher education upon graduation. These two tracks combined contain approximately 45 percent of all Dutch secondary school students.

In the first three years of HAVO and VWO, all courses are mandatory.

Schools are free to allocate hours between courses within reasonable bounds. For example, one school may choose to give their students three hours of math per week, while other schools might give them four hours. After the first three years, students have to choose one of four fields of specialization. These four fields are *Natuur en Techniek* (Nature and Technology), *Natuur en Gezondheid* (Nature and Health), *Economie en Maatschappij* (Economics and Society) and *Cultuur en Maatschappij* (Culture and Society). The field is important for the major choice in tertiary education. For example, only the field Nature and Technology (*Nature/Tech*) gives access to all bachelor fields and is the only field that automatically gives access to all STEM bachelors. The other fields only give access to a subset of bachelors.

Each field consists of a combination of three or four subjects that are mandatory for the specific field of specialization. For Nature/Tech, these subjects are Physics, Chemistry and Mathematics B, which has the most intensive and challenging math curriculum. Beside the field part, the second stage consists of compulsory subjects like Dutch, English and physical education that have to be followed regardless of the field choice. Moreover, there is an elective part where students are required to take one or two extra electives which can be either related to the field choice or not. At the end of the second stage, all students write a centralized exam.

Due to the electives, it is possible for a student to meet the criteria to graduate in two or more fields. Officially, you can graduate with two fields of specialization. Most commonly, these combinations are the similar fields "Nature and Technology" and "Nature and Health" and "Economics and Society" and "Culture and Society". In this paper, we consider someone who has graduated high school with the subjects Physics, Chemistry and Mathematics B as Nature/Tech graduates since these students will have access to all (STEM) bachelors.

1.2.2 Tertiary education in the Netherlands

After the students pass the final examinations, they leave secondary education. Both the HAVO and the VWO track satisfy the requirements of the Dutch compulsory education law. This means that HAVO and VWO graduates can either leave the education system, go to secondary vocational education (MBO) or pursue a Bachelor's degree. HAVO graduates can go to a university of applied sciences (HBO) and graduates with a VWO degree can go to either a research university (WO) or HBO. A bachelor in HBO takes four years and a bachelor in WO takes three years. Therefore, the years of education from the start of secondary school to a Bachelor's degree are equal to nine years for students from both the HAVO and VWO tracks. Students have to pick a major as soon as they enroll in a WO or an HBO.

When enrolling for a major in a WO or HBO, the first admission criterion is the field of specialization in secondary school. As mentioned before, graduating secondary school with a Nature/Tech specialization gives access to all bachelors. A bare pass suffices for all bachelors without quota. The other three fields give access to only a subset of bachelors. The focus of this paper lies in the choice of a STEM major. In this paper we rely the definition of the commonly used International standard classication of education ISCED (2011) according to which categories 4 and 5 are considered to be STEM majors. The majors belonging to these groups are displayed in Table A.1 in the Appendix.

It is important to note that in order to be able to enroll in all of the STEM majors in Table A.1, the student needs to have a degree with the Nature/Tech field. This means that if a student at the end of their third year of secondary education (at age 15) decides not to choose the Nature/Tech field, they will lose the opportunity to obtain STEM degrees in university or HBO. This makes the field choice in high school a high-stakes decision with consequences for the long run.

The secondary school grades are only important for bachelor studies with a quota. This applies for example to medicine for research universities or physiotherapy for universities of applied sciences. For some majors, there are quota at certain universities but not at others (e.g. business administration). When selecting students, universities would generally value a Nature/Tech degree more than a degree in one of the other fields. However, a selection committee for a major like Business Administration might value a prospective student with an Economics and Society degree with perfect grades more than one with a bare pass for Nature/Tech.

Due to the large demand for STEM graduates, there are barely any quota for STEM majors. The only STEM major which had cap limit on admissions for the treated group was Clinical Technology at the University of Twente DUO (2011). However, the enrollment that year did not exceed the quota. So everybody with a Nature/Tech degree who wanted to do a STEM field, could enroll in the major at the university of their first choice.

Table 1.1: Policy change in 2007: Changes in Hours of Education

	HAVO			VWO		
	Old	New	Δ	Old	New	Δ
Compulsory	1,480	1,120	-360 (-24.3%)	1,960	1,920	-40 (-2,0%)
Field Courses	1,160	1,080	-80 (-6.9%)	1,840	1,520	-320 (-17.4%)
Electives	560	1,000	440 (+78.6%)	1,000	1,360	360 (+36%)
Total	3,200	3,200	0 (0%)	4,800	4,800	0 (0%)

1.2.3 Policy change

In August 2007, changes were applied to the second stage of secondary school, i.e. after students' decision on the field of specialization.[4] One of the advocates of this policy reform was the Platform Beta Techniek, which was founded in 2004. This platform was a collaboration of the Dutch ministries of Education, Economic Affairs and Social Affairs and Employment which had a goal to increase the amount of (female) STEM graduates in (research) universities.

The main goals of this policy change were threefold. First of all, the ministry of Education wanted to simplify the structure of the second stage. Secondly, the new second stage was supposed to give students more freedom in choosing their curriculum by allotting extra time to electives (du Pre, 2005). Thirdly, this policy change was specifically aimed at increasing the fraction of people choosing and completing the *Nature/Tech* field, especially among women. In the cohorts before the policy change, less 10 percent of female students graduated secondary school with the *Nature/Tech* field.

A summary of the main policy changes is displayed in Table 1.1, while a more detailed description of the policy change can be found in Table A.2 in the Appendix. In VWO, the number of course hours for the field part decreased by 17 percent (from 1,840 to 1,520 hours). In HAVO, the field part decreased by seven percent (from 1,160 to 1,080 hours). For VWO, this implies that –per week– two hours and 40 minutes less are spend on the three STEM subjects (math, physics and chemistry). For HAVO, the reduction amounts to only one hour. The second (field) stage was subject to several other changes beside the reduction in STEM hours. For example, the number of hours for electives increased and the number of hours for courses without final exam (such as P.E.) decreased. However, all other changes in

[4]The source for the policy change is the document *Zakboek Tweede Fase* (2007) written by the institution Tweede Fase Adviespunt which advised the Dutch ministry of Education until 2009.

the second stage were identical for both HAVO and VWO. Therefore, any differential change in field choices in HAVO and VWO should be due to the differential changes in course hours in the field part.

This policy change thereby lowered the bar to obtain the most-valued degree in secondary schools, which allows to enroll in any major in college/university. At the same time, the number of field-specific hours (such as math, physics and chemistry in the STEM field) were reduced (while the total instruction time stayed constant because the reduction was compensated by an increase in hours in elective subjects). Since hours of instruction are a good proxy for the knowledge acquired in a course,[5] lowering the number of mandatory field hours therefore reduced the STEM content in *Nature/Tech* without the option to have the same level of STEM preparation as before the policy change.

The policy change went into effect on August 1, 2007 and applied to everyone entering the fourth year of HAVO or VWO starting with the academic year 2007-2008. This policy did not affect students in the higher years in 2007. This means all HAVO and VWO students born after October 1, 1991 should be affected by the policy change if they did not skip a class. All students born before that date are not affected by the policy change and completed their degree with the old second stage (unless they had to repeat a year).[6]

The first question we address is what are the short-run effects of the policy change. There is a cost reduction of choosing *Nature/Tech* field in high school, which could affect the preferences for choosing the *Nature/Tech* field. As one of the goals of the policy change was to increase the number of *Nature/Tech* graduates among underrepresented groups, the question is how the effect of the policy change differed by gender, socio-economic and migration background.

Secondly, we analyze the longer-run effects of the policy change in terms of college majors. As the policy change made *Nature/Tech* more accessible, the question is how the enrollment and graduation change in terms of STEM bachelors and STEM masters changed and whether this effect differs by gender, socio-economic and migration background. Moreover, as we can follow individuals until they are 29, we can study the longer-run effects in terms of earnings and family formation outcomes.

[5] See, for example, Lavy (2020) and Lavy (2015), who show that hours of instruction have a positive effect on academic performance.

[6] Using individuals' birth year is the conservative way to classify students to avoid selecting into or out of the policy change.

1.3 The empirical model

In the empirical analysis, we exploit the policy change of 2007, which reduced field-specific hours in the VWO (academic) relative to the HAVO (applied academic) track. Thereby, the policy change reduced the study load and effort costs of the STEM field. In particular, students saw a reduction of 17.4 percent in the number of field-specific course hours in VWO compared to a reduction of 6.9 percent in HAVO. This means that the reform makes it easier to meet the prerequisites for a STEM major (i.e. the completion of the Nature/Tech field in secondary school) in VWO compared to HAVO.

We make use of the policy change as a quasi-experiment where we consider VWO as the treatment group (large reduction of field specific hours) and HAVO as control group (small reduction of field-specific hours) to control for counterfactual trends, i.e. for how decisions would have changed in the absence of the reform. To identify the total effect of the change in field-specific hours on choices, it would have been ideal if there had been no corresponding change in the HAVO track. However, evaluating the effect of a large reduction of field-specific hours (in VWO) compared to a smaller reduction (in HAVO) should lead to conservative estimates of the effects and provide us with lower bounds of the true effect, as discussed further below.

We use a difference-in-differences (DID) setup to exploit the differential reduction of field-specific hours for the two different tracks. Our main model can be seen in equation 1.1, where y_{itc} denotes the outcome of interest, including the choice of the field/specialization *Nature/Tech* as well as a number of longerrun outcomes. VWO_{it} is an indicator for whether student i is in the VWO (1) track or in the HAVO (0) track. LC_{ic} is an indicator which is 1 if student i is born in the later-born cohort affected by the policy change and 0 if student i is born in the earlier-born cohort unaffected by the policy change. $Treatment_{itc}$ is an indicator which takes the value 1 if a student is in the later-born treated cohort and in the VWO (i.e. the treated) track. X_{itc} includes a set of controls including gender, migration background, parental background and municipality of residence. β_1 denotes the coefficient of interest and shows the causal effect of the policy reform.

$$y_{itc} = \beta_0 + \beta_1 * Treatment_{itc} + \beta_2 * VWO_{it} + \beta_3 * LC_{ic} + X_{itc} * \gamma + \epsilon_{itc} \quad (1.1)$$

As increasing the representation of underrepresented groups in STEM fields was one of the goals of the policy change, it is important to investigate whether and how the effects of the reform differed for different subgroups. We therefore also estimate a fully interacted model with a group indicator G (see equation 1.2), where the main coefficient of interest δ_1 can be interpreted as a differential treatment effect. When the interaction effects show a significantly different response to the reform for a particular subgroup, we estimate the non-interacted DID model 1.1 on the relevant subgroup only. We look at gender, migration background and socio-economic status (household income) as our subgroups of interest for our heterogeneity analysis.

$$y_{ist} = \beta_0 + \beta_1 * Treatment_{itc} + \beta_2 * LC_{ic} + \beta_3 * VWO_{it} + \\ \delta_1 * Treatment_{itc} * G_{itc} + \delta_2 * LC_{ic} * G_{itc} + \delta_3 * VWO_{it} * G_{itc} + X_{itc} * \gamma + \epsilon_{itc} \tag{1.2}$$

1.4 Data

For this study, we make use of results based on our own estimations and calculations using the non-public administrative micro-database from Statistics Netherlands (*Centraal Bureau voor de Statistiek, CBS*). This database contains information on the entire Dutch population and are particularly suitable for our purpose, since they contain information on month and year of birth, educational histories including track in secondary school (such as HAVO, VWO), highest completed degree (such as bachelor, master), fields in secondary school (such as Nature/Tech) and majors at college/university (such as STEM), yearly income, individual and household characteristics, such as gender, socio-economic status of the family, parental education and occupation (including STEM background) and migration background.

In our analysis we compare two cohorts of individuals, a younger cohort which was affected by the reform and an older cohort that was too old to be affected by the reform. More specifically, the cohort born between 1 October 1990 and 30 September 1991 was aged 16 and attending grade 11 at the time of the reform in 2007. The reform did not apply to this cohort, as they had already chosen their field and the number of field-specific hours remained at the level prior to the reform until they completed secondary education. Instead, the cohort born one year later (between 1 October 1991 and 30 September 1992), were aged 15 and attending grade 10. Thus, they were the first cohort to whom the reform applied. They started and completed the field given the new rules. When they had to choose their field (i.e. *Nature/Tech* or one of the other three), the new cohort was aware of the change in field-specific hours as the law passed in April 2006 (Wijzigingswet Voortgezet Onderwijs, 2006), more than a year before they had to make their field choice.

Importantly, for all students in these cohorts we know their track allocation, their field choice, their centralized exam scores, their tertiary education enrollments and degrees until 2021 and their income until 2020. In the following sections, we describe the definition and construction of the variables we use and present descriptive statistics.

1.4.1 Construction of variables

In this section, we describe the variables we use from the CBS microdatabase for our analysis as well as how we constructed these variables. The main dependent variables we are looking at are related to short-run and long-run educational attainment. In terms of short-run outcomes, we are interested in whether the STEM field *Nature/Tech* is chosen or not. *Nature/Tech* is a dummy variable that is 1 if a student completed secondary school with the *Nature/Tech* field and 0 otherwise. *STEM Bachelor* and *STEM Master* take the value of 1 if a student got a Bachelor and Master in one of the STEM fields, respectively, and 0 otherwise.

For the longer-run outcomes, *Graduation delay* is the number of months that a student needed to graduate on top of the expected time of completion. *Personal income* is the logarithmic personal income of an individual at age 28. *Partner* takes the value of 1 if an individual has a registered partner or a spouse by age 29, and 0 otherwise. *Married* takes the value of 1 if an individual has a spouse by age 29, and 0 otherwise. *Child(ren)* takes the value of 1 if an individual has at least one child by age 29, and 0 otherwise.

We analyze the heterogeneity of effects of the policy along three dimensions, gender, socio-economic status and migration background. The corresponding variables are defined as follows. *Female* is 1 if a student is female. *Migration Background* is 1 if the student or at least one of their parents are born outside of the Netherlands.[7] Finally, we categorize the cohorts by household income in the year when the students are choosing their high school field. *Low Income* takes the value 1 if a student is from a low-income household. Since students in the two highest tracks in secondary school are from the more privileged part of society, only around 20

[7] We also considered alternative definitions, such as students with a non-Western migration background or students with at least both parents born abroad, which led to similar results.

percent of students have a household income below the 60th percentile. To have a sufficient number of households in the Low Income category, while the variable should still capture coming from a less privileged background, we use the 60th percentile based on which we define a Low or High Income background, but alternative definitions generate similar results (as discussed further below).

1.4.2 Descriptive Statistics

We present descriptive statistics of the main variables used in our analysis in Table 1.2. As expected, there are some differences between students in the VWO (academic) and HAVO (applied academic) track. Students in the academic track of high school are somewhat more likely to, among others, be female (54 versus 51%), come from a two parents household (77 versus 72%), come from a higher income household (percentile 78 versus 72) and have at least one parent with a higher education degree (34 versus 24%). However, the differences within tracks and between cohorts are small and mostly insignificant.

1.5 Short- und Long-Run Effects of the Reduction in STEM Hours

1.5.1 Short-run effects: *Nature/Tech* graduation

In this section, we show the effect of the policy on graduating secondary education with the *Nature/Tech* field. We first show the results graphically and then present regression results in tables. In our regression analysis we investigate the pooled effects as well as effects separately by gender and household income to see if the policy change positively affected students from underrepresented groups.[8]

One can see the immediate effect of the policy in Figure 1.1. The policy change had the expected shortrun effect. The number of students choosing *Nature/Tech* increased significantly more in VWO (academic track) compared to HAVO (applied academic track). More specifically, the number of *Nature/Tech* graduates in VWO increased by 14.7 percentage points (from 17.3 percent in the earlier cohort still subject to the old rules to 32 percent among the younger cohort subject to the reduction in field-specific hours). Meanwhile in HAVO, there is an increase in the take-up of *Nature/Tech*,

[8]We also analyze effects by migration background, but we do not find differential responses to the policy change (see Table A.4 in the Appendix).

Table 1.2: Population Summary Statistics

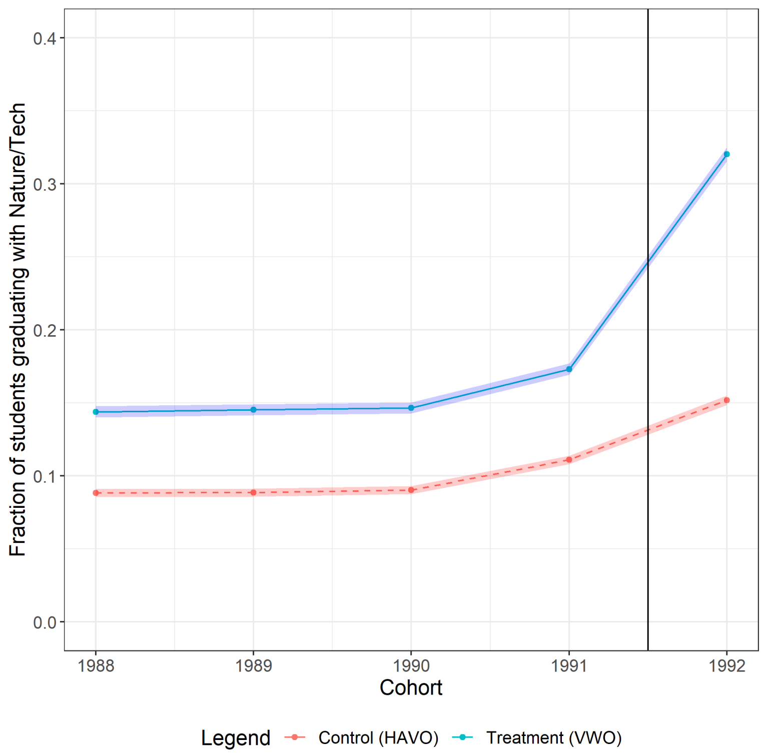
	VWO			HAVO		
	Younger Mean (SD)	Older Mean (SD)	Diff (p-value)	Younger Mean (SD)	Older Mean (SD)	Diff (p-value)
Female	.543	.542	.001	.514	.518	−.004
	(.498)	(.498)	(.731)	(.5)	(.5)	(.305)
Birth Month	6.356	6.368	−.012	6.532	6.566	−.034
	(3.47)	(3.469)	(.612)	(3.45)	(3.45)	(.131)
Siblings	1.682	1.694	−.012	1.76	1.77	−.01
	(1.098)	(1.101)	(.051)	(1.24)	(1.222)	(.189)
Two Parents HH	.772	.767	.005	.723	.717	−.006
	(.419)	(.423)	(.128)	(.448)	(.45)	(.049)
Migration Background	.170	.172	−.002	.176	.174	.002
	(.376)	(.377)	(.397)	(.381)	(.379)	(.485)
Non Western Migration	.091	.093	−.002	.114	.109	.005
	(.287)	(.291)	(.258)	(.318)	(.312)	(.019)
Both Parents Foreign	.079	.081	−.002	.100	.099	.001
	(.27)	(.274)	(.313)	(.264)	(.298)	(.283)
Observations	38,191	37,625		46,264	45,701	
HH Income Percentile	78.26	78.04	.22	73.07	73.34	−.27
	(20.449)	(20.64)	(.289)	(21.79)	(21.6)	(.118)
Low Income	.161	.165	−.004	.229	.224	.005
	(.368)	(.372)	(.107)	(.42)	(.417)	(.100)
Observations	37,242	36,660		45,043	44,518	
Parent with College	.347	.331	.016	.239	.235	.004
	(.476)	(.47)	(<.001)	(.426)	(.424)	(.323)
Parent with STEM	.114	.114	<.001	.075	.072	.003
	(.318)	(.318)	(.998)	(.258)	(.264)	(.111)
Observations	26,701	25,723		31,220	30,176	

Note: HH: Household

albeit to a much smaller extent (increase of 4.1 percentage points from 11.1 to 15.2 percent). As we use the individuals' birth year as the conservative way to classify students to avoid selection, the increase starts one birth cohort earlier. This is due to some students in the earlier birth cohort making the choice under the new set of rules.

To analyze the overall effect of the policy on the likelihood of completing a *Nature/Tech* degree, we estimate equation 1.1 on the pooled sample. More

Figure 1.1: Fraction of the students graduating high school with Nature/Tech by birth cohort



Note: The y-axis displays the fraction of students who graduated with the Nature/Tech field in high school. The x-axis displays the birth cohort. The policy change only applies to those born in birth cohort 1992. However, due to retainers some students born in birth cohort 1991 also chose a field of graduation after the policy change. The shaded areas indicate 95 percent confidence intervals.

specifically, we analyze how the likelihood of completing the *Nature/Tech* field changed post compared to prior to the reform in the VWO (academic) track in which students experienced the drastic reduction in field-specific hours, using as the counterfactual trend the change in the likelihood of *Nature/Tech* in the HAVO (applied academic) track, where students only experienced a minor decrease in field-specific hours. We present results from specifications without and with controls. To provide supporting evidence for the parallel trend assumption, we also present results from a placebo test where we test whether pre-reform trends (for two earlier cohorts) were indeed parallel for the two tracks. In the main tables we only present the main coefficients on the treatment indicator, but present means for the four groups (the two tracks and the cohorts) in Table A.3 in the Appendix.

The short-run effect of the policy change can be found in the Table 1.3. In Panel A, the effect is shown for all students. The reduction in STEM hours leads to an increase in the likelihood of graduating secondary school

Table 1.3: Short run effects: Graduating secondary school with Nature/Tech

	Nature/Tech Field Completion			
	Main		Placebo	
	No controls	Controls	No controls	Controls
	(1)	(2)	(3)	(4)
Panel A: All students				
Treatment	.108***	.108***	−.0005	−.001
	(.004)	(.004)	(.003)	(.003)
Observations	154,042	154,042	141,719	141,719
Panel B: By gender				
Treatment **for women**	.137***	.137***	−.0004	−.001
	(.004)	(.004)	(.003)	(.003)
Observations	81,340	81,340	74,890	74,890
Treatment **for men**	.075***	.074***	−.001	−.002
	(.004)	(.004)	(.006)	(.006)
Observations	72,702	72,702	66,829	66,829
p-value of the difference		<.0001		.978
Panel C: By household income				
Treatment **for low income households**	.070***	.074***	−.004	−.004
	(.009)	(.009)	(.007)	(.007)
Observations	30,211	30,211	28,530	28,530
Treatment **for high income households**	.114***	.115***	−.0004	−.0004
	(.004)	(.004)	(.004)	(.004)
Observations	123,831	123,831	113,189	113,189
p-value of the difference		<.0001		.617
Control variables	NO	YES	NO	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. In columns 1 and 2, the cohorts of interest are analyzed. In columns 3 and 4, the two cohorts before the cohorts of interest are analyzed. The regressions in the odd columns only include the three difference-in-differences indicators, while the regressions in the even columns include control variables on the individual level.

with a *Nature/Tech* degree by 10.8 percentage points.

As women are underrepresented in STEM fields, it is interesting to see if the policy change affected women differently than men. In Panel B, we run the regression from equation 1.1 separately by gender and we use a triple interaction from equation 1.2 to measure if the causal effect of the policy change are statistically differed by gender. There is a significant difference between how males and females were affected by the policy change. Both genders have more *Nature/Tech* graduates after the policy change, but women are significantly more likely to be affected than man.

In Panel C, it can be seen that both students from high and low income households are positively affected by the policy, but students from high income households increase their likelihood of graduating with a *Nature/Tech* degree more strongly than students from low income households.

Thus, in the short-run, the policy change led to a significant increase in the number of students meeting the prerequisites for a STEM major at college/university, overall and for each of the subgroups. Especially women are more likely to meet the prerequisites to obtain a STEM major. However, the gap between high and low SES students obtaining a *Nature/Tech* degree increased.

1.5.2 Longer-run effects: Tertiary STEM education

The effects on graduating tertiary education with a STEM degree can be seen in Table 1.4. The variables of interest are having a STEM bachelor degree and a STEM master's degree. We have also considered the enrollment rates instead of completion rates, but the effects on enrollments can be found in Table A.5 the Appendix. The completion rates are the main outcomes as they have more severe consequences.

It can be seen in Panel A that the DID-coefficients for the full population are not statistically different from zero. This means that there is no increase in the fraction of people graduating in a STEM major after the policy change. This is a remarkable result, as the group with the prerequisites to enroll in the STEM majors is 10.8 percentage points larger. It can be seen in Panel B that men are actually more likely to graduate with the STEM degree, but women are not more likely to be graduating with a STEM major. In fact, women are less likely to graduate with a STEM major than men. This is surprising, as women are significantly more likely to be graduating with *Nature/Tech* because of the policy change compared to men, but significantly less likely to graduate with a STEM major compared to men. This means that the policy change kept the number of STEM graduates approximately equal, but the composition of STEM graduates has a higher share of males.

Low SES students are less likely to be graduating with a STEM master. Moreover, they are less likely to graduate with a STEM bachelor or STEM master compared to students with high SES. High SES students are more likely to graduate with a STEM bachelor. This means that the policy change leads to more high SES and male graduates, even when the total number of STEM graduates did not change significantly.

Table 1.4: Tertiary education effects: Graduating tertiary education with a STEM degree

	STEM Degree Completion			
	STEM Bachelor		STEM Master	
	Main	Placebo	Main	Placebo
	(1)	(2)	(3)	(4)
Panel A: All students				
Treatment	.005	.0004	.002	−.003
	(.003)	(.003)	(.003)	(.003)
Observations	154,042	141,719	154,042	141,719
Panel B: By gender				
Treatment **for women**	-.003	.001	−.005	.0003
	(.004)	(.004)	(.004)	(.004)
Observations	81,340	74,890	81,340	74,890
Treatment **for men**	.014***	−.0003	.010**	−.006
	(.005)	(.005)	(.004)	(.004)
Observations	72,702	66,829	72,702	66,829
p-value of the difference	.004	.960	.002	.318
Panel C: By household income				
Treatment **for low income households**	−.007	−.001	−.012*	−.004
	(.007)	(.007)	(.006)	(.007)
Observations	30,211	28,530	30,211	28,530
Treatment **for high income households**	.007**	.001	.005	−.002
	(.003)	(.003)	(.003)	(.003)
Observations	123,831	113,189	123,831	113,189
p-value of the difference	.080	.803	.010	.775
Control variables	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In columns 1 and 2, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In columns 3 and 4, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. In the odd columns, the cohorts of interest are analyzed. In the even columns, the two cohorts before the cohorts of interest are compared as a test for pretrends. All regressions include control variables on the individual level.

1.6 Mechanisms

Why is there is no absolute increase in the number of female STEM graduates, even if there are 13.7 percentage points more women with the prerequisites to do a STEM major? One possibility is that all women who were taking *Nature/Tech* only because of the policy change simply graduated in the non-STEM field as they would without the policy change.

It is impossible that the non-increase of female STEM graduates is due to any admission caps. As there are no STEM majors which had to decline any applicants. First, we show that women are actually less likely to graduate with a STEM bachelor and STEM master if they have at least one parent with a STEM college degree. Then we show that this does not hold for women with a parent who has a college degree in other fields.

1.6.1 Parental education

To show that women are affected by the policy change too, we look at the educational background of the parents. Apparently, household income has an influence on the outcomes of the individuals. The question we now address is whether the parental education does too. The problem with using parental education is the fact that we can not perfectly link the education of the parents to the children with the administrative data. If we include parental education in our set of variables, around 32.5 percent of the observations go missing. The missing observations are also the reason why we use household income as indicator for SES in the main analysis. In this section, we do include parental education as a subgroup.

Even though it is not realistic to assume that the 32.5 percent observations without information on the parental education are missing completely at random, it does seem that the subgroup of individuals from whom we observe the parental education responds similary to the policy change as the full sample. In Table 1.5, we show the treatment effect for people with and without STEM parents. *STEM parents* is 1 if at least one parent has a degree in higher education in a STEM field.

Students who have a STEM parent are 7.3 percentage points more likely to get the *Nature/Tech* degree after the policy change as can be seen in Panel A. In Panel B1, it can be seen that women in the treated cohort who have at least one parent with a STEM degree are 7.5 percentage points more likely to get the *Nature/Tech* degree in high school than women without a STEM parent. However, women are 3.8 percentage points less likely to get a STEM bachelor and 2.4 percentage points less likely to get a STEM master in tertiary education compared to the women in the treated cohort who do not have at least one parent with STEM degree. In Panel B2, it is shown that men are 5.9 percentage points more likely graduate with *Nature/Tech*, but they are not affected in their likelihood to get a STEM degree if they have a parent with a STEM degree.

If women would not change their major choice at all, we should not have observed differences between different levels of parental education. To see whether the decrease of STEM degrees for women with STEM parents is due

Table 1.5: Parents with a STEM degree

	Main Cohorts		
	Nature/Tech	STEM Bachelor	STEM Master
	(1)	(2)	(3)
Panel A: All students			
Treatment x STEM parents	.073***	−.010	−.012
	(.018)	(.012)	(.011)
Treatment	.099***	.004	.003
	(.005)	(.004)	(.003)
Observations	104,879	104,879	104,879
Panel B1: Women			
Treatment x STEM parents	.075***	−.038***	−.024*
	(.021)	(.017)	(.015)
Treatment	.128***	−.002	−.005
	(.005)	(.005)	(.005)
Observations	55,265	55,265	55,265
Panel B2: Men			
Treatment x STEM parents	.059**	.018	.002
	(.029)	(.018)	(.016)
Treatment	.066***	.012**	.011**
	(.008)	(.006)	(.005)
Observations	49,614	49,614	49,614
Control variables	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field. In column 2, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In column 3, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x STEM parents is a triple interaction term that is 1 if an individual is treated and has at least one parent with a college degree in a STEM field. All regressions include control variables on the individual level.

to information they receive from the parents, we can compare these results to parents with another college degree. In Table 1.6, the interaction effect of the treatment with a parent having a college degree is shown. *College parents* is 1 if at least one parent has a degree in higher education.

Table 1.6 shows that neither women nor men with a college parent are less likely to graduate with a STEM bachelor or master. This means that women with a STEM parent, unlike men with a STEM parent, are unconditionally less likely to do a STEM major themselves after the policy change. This holds even when these treated women are even more likely to

Table 1.6: Parents with a College degree

	Main Cohorts		
	Nature/Tech	STEM Bachelor	STEM Master
	(1)	(2)	(3)
Panel A: All students			
Treatment x College parents	.046***	.002	.004
	(.011)	(.008)	(.007)
Treatment	.092***	.002	−.001
	(.006)	(.004)	(.004)
Observations	104,879	104,879	104,879
Panel B1: Women			
Treatment x College parents	.042***	−.004	−.007
	(.012)	(.011)	(.010)
Treatment	.123***	−.006	−.006
	(.006)	(.006)	(.005)
Observations	55,265	55,265	55,265
Panel B2: Men			
Treatment x College parents	.051***	.008	.015
	(.018)	(.012)	(.011)
Treatment	.056***	.011	.006
	(.010)	(.007)	(.006)
Observations	49,614	49,614	49,614
Control variables	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field. In column 2, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In column 3, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x College parents is a triple interaction term that is 1 if an individual is treated and has at least one parent with a college degree. All regressions include control variables on the individual level.

do *Nature/Tech* than treated women without a STEM parent.

This could be explained as follows: Women who would not have done *Nature/Tech*, can now do it at a lower cost as one needs to master less STEM content to receive the highest high school degree. However, the parents of the child with the experience of a STEM major, might signal that, from their experience, the child misses crucial abilities to get a STEM degree and this might increase the perceived costs of the women to get this degree.

The effect is not driven by having highly educated parents in general, as women with a parent with a college degree another field are not affected by the policy change. That men are not affected can be explained by the fact that men have higher self-confidence (Morin (2015), Preckel et al. (2008), Niederle and Versterlund (2007)), and are less sensitive to their surroundings when choosing a major (Mouganie and Wang (2019), Cools et al. (2019)).

1.6.2 Longer run benefits from choosing *Nature/Tech*

The question that remains is if it was beneficial for women, especially women with a STEM parent, to choose *Nature/Tech* in high school, despite not doing more STEM majors. One reason for women to opt for another major could be advantages after tertiary education. Therefore, we look at the longer run effects of the policy change. We can observe the treated cohort until they are in their late twenties (age 29). We can look at the effect of the policy change on their income, how much study delay they had (how many months did students exceed nominal duration of their studies), whether they had a fiscal partner and the characteristics of the partner, whether they had a spouse and the characteristics of this spouse, and whether they had children and the age of first time parenthood.

The longer run effects could help to understand why some women opt out of the STEM fields. Women who do a STEM field in high school, might have opted for a non-STEM major as they might believe they could get higher earnings or meet partners with higher quality.

Table 1.7 shows the estimations of the β_1 coefficients for the longer run variables of interest. Again, we look at the results for the subsamples split by gender and household income.

There are some longer run differences from the policy change. In the treatment group, people are around 2.5 percentage points more likely to have a partner with *Nature/Tech*. Males are slightly more likely to find a partner with a STEM Bachelor. This seems paradoxical at first, as women are less likely to do a STEM Bachelor after the policy change. However, we also find that treated men are more likely to get a partner with a VWO degree. As the absolute number of female VWO students with a STEM bachelor is higher than for female HAVO or VMBO students, it makes sense that the treated men are more likely to have a partner with a STEM bachelor. There seem to be no positive longer run effects for women from the policy change, beside the higher likelihood for a women to have a partner with a *Nature/Tech* degree.

Table 1.7: Longer run Treatment effect by subgroup

	Treatment			
	Females	Males	Low Income	High Income
	(1)	(2)	(3)	(4)
Graduation delay (months)	.500	−.317	−.108	.139
	(.350)	(.396)	(.641)	(.289)
Bachelor	−.001	.001	.007	−.003
	(.005)	(.006)	(.010)	(.004)
Master	.010	.008	−.010	.013***
	(.006)	(.006)	(.010)	(.005)
Personal Income	−.012	.016	−.007	.004
	(.021)	(.025)	(.043)	(.017)
Partner	.006	.007	.010	.006
	(.005)	(.006)	(.010)	(.004)
Married	.002	.005	.009	.002
	(.006)	(.005)	(.009)	(.005)
Child(ren)	−.007	.001	.006	−.004
	(.006)	(.005)	(.009)	(.004)
Partner Personal Income	.011	−.059*	.021	−.003
	(.021)	(.030)	(.047)	(.004)
Partner Nature/Tech	.020***	.031***	.020***	.027***
	(.004)	(.005)	(.007)	(.003)
Partner STEM Bachelor	.0002	.017**	.005	.008
	(.007)	(.007)	(.011)	(.006)
Partner STEM Master	.001	.005	.004	.002
	(.003)	(.003)	(.005)	(.002)
Observations	81,335	72,739	30,222	123,852
Controls	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. Columns 1 and 2 show the effect of the treatment on women and men, respectively. Columns 3 and 4 show the effect of the treatment on students from low- and high income households, respectively. The rows indicate the different dependent variables of interest. The coefficients show the point estimate of β_1, the DID estimator. All regressions include control variables on the individual level.

1.6.3 Longer run benefits for women with STEM parents

As women with a STEM parent are even less likely to get a STEM major, the question is why these women opted for another major. Did they make the right choice or not?

Table 1.8: Longer run effects for women with parental STEM education

	Parental Background Women			Parental Background Men		
	STEM	no STEM	Diff	STEM	no STEM	Diff
	(1)	(2)	(3)	(4)	(5)	(6)
Income	.218**	−.042	.253**	−.061	−.005	−.073
	(.093)	(.027)	(.094)	(.114)	(.033)	(.119)
Married	.050*	.002	.054**	.010	.003	.011
	(.026)	(.026)	(.027)	(.021)	(.007)	(.023)
Children	.037	−.013*	.047*	.019	.001	.013
	(.025)	(.008)	(.026)	(.019)	(.006)	(.020)
Observations	4,329	48,764	53,093	4,259	43,311	47,570
Control variables	YES	YES	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is the logarithmic gross income at age 29. In column 2, the dependent variable is an indicator variable which is 1 if the individual got married by age 29 and 0 otherwise. In column 3, the dependent variable is an indicator variable which is 1 if the student had at least one child by age 29 and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x STEM parents is 1 if an individual is treated and has at least one parent with a college degree in a STEM field. All regressions include control variables on the individual level.

From Table 1.8, it can be seen that the policy change influenced the women with a STEM parent in other ways than only decreasing the likelihood of graduating with a STEM major. In fact, women with a STEM parent have higher income, are 5.4 percentage points more likely to have a spouse and are 4.7 percentage points more likely to have at least one child by age 29. The same is not true for men with a STEM parent. They are not affected through their parents' education by the policy change, apart from being more likely to do *Nature/Tech*.

1.6.4 Classroom composition

The final question is what the reason for women and students with high SES was to choose *Nature/Tech* after the policy change. One reason is the option value of having the *Nature/Tech* degree. By choosing *Nature/Tech*, you are able to choose any major. After the policy change, it was easier to get this degree. A second reason is students having a more quantitative background when they are leaving secondary school. This could be helpful for certain non-STEM majors and as a way to signal your value to future employers. This is more attractive after the policy change, as the *Nature/Tech* degree can be obtained at a lower cost.

Table 1.9: Classroom composition

	Nature/Tech			
	Females	Males	Low Income	High Income
	(1)	(2)	(3)	(4)
Panel A: More women in Nature/Tech				
Treatment x Share of Women	.002	.003	−.011	.004
	(.009)	(.015)	(.019)	(.009)
Treatment	.137***	.070***	.078***	.112***
	(.007)	(.011)	(.014)	(.007)
Observations	77,976	70,213	29,083	119,106
Panel B: More high income students in Nature/Tech				
Treatment x Share of High Income	.022***	.005	−.025	.016*
	(.009)	(.014)	(.018)	(.009)
Treatment	.128***	.074***	.087***	.109***
	(.006)	(.010)	(.013)	(.006)
Observations	77,976	70,213	29,083	119,106
Control variables	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field. Columns 1 and 2 show the effect of the treatment on women and men, respectively. Columns 3 and 4 show the effect of the treatment on students from low- and high income households, respectively. Treatment shows the point estimate of β_1, the DID estimator. Treatment x Share of Women is 1 if an individual is treated and graduated from a secondary school that had a higher than median predetermined share of female students graduating with the Nature/Tech field. Treatment x Share of High Income is 1 if an individual is treated and graduated from a secondary school that had a higher than median predetermined share of high income students graduating with the Nature/Tech field. All regressions include control variables on the individual level.

There is also the option of networking. By choosing *Nature/Tech*, you share STEM courses with your peers in your school who also chose *Nature/Tech*. We look at the effect of the policy on the different subgroups regarding the predetermined share of females and high SES students in the *Nature/Tech* field in Table 1.9. We take the fraction of women and high SES students in *Nature/Tech* in a certain school in the cohort before the cohorts of interest to prevent endogeneity.

It is shown in Panel A that the share of women in *Nature/Tech* does not matter for any subgroup when choosing *Nature/Tech* after the policy change. In Panel B, it can be seen that the predetermined share of high income students is important. The treatment effect is amplified for women and students from a high income households. A school with an above median predetermined share of high income students with *Nature/Tech* leads to 2.2 percentage points more women and 1.6 percentage points more students from high income households choosing *Nature/Tech* after the policy change.

This gives reason to suspect that there is a networking effect for women and high income students. Women encounter more potential high income and highly educated marriage market candidates by choosing *Nature/Tech*. High income students can benefit from networking with other high income students by choosing *Nature/Tech* in schools that are known to have a higher share of high income students with *Nature/Tech*.

1.7 Concluding remarks

This paper examined the effect of a decrease in the amount of STEM hours in high school which are a prerequisite for doing a STEM major in tertiary education. The policy does increase the amount of eligible students substantially. There is a much larger fraction of people that will choose the STEM field in high school with the new STEM requirements. These students benefit from the lower requirements because they finally consider it worth the cost to pursue the STEM field in high school which is the strongest degree when applying for a major in tertiary education. However, there is no evidence that the government was able to increase the number of STEM graduates by lowering the STEM requirements in high school.

Another important finding is that there are heterogeneous effects of this policy change. More females are persuaded in graduating secondary school with a *Nature/Tech* degree after this policy change than males do. However, fewer females will actually complete a STEM bachelor compared to men because of the policy change.

This result might seem paradoxical. However, it follows the idea that women require a stronger signal in order to be confident about their abilities (Justman and Méndez, 2018). When secondary school students get less exposure to STEM, females might not be confident enough to pursue a STEM major compared to if they did have more exposure to STEM.

If we only look at women, there is even more heterogeneity when we separate according to parental education. Women with at least one parent with a STEM degree are the ones that are less likely to do a STEM major. Nevertheless, the parental background does not affect the effect of the policy change on men. This supports the idea that this reduction of women in STEM majors is due to a problem in different expectations, as women with a STEM parent would have a good role model and should have good opportunities to do a STEM major in general. These women do have positive effects at the end of their twenties as they have a higher salary and are more likely to be married and have children.

Students from lower income households are both less likely to be affected by the policy in the short run and less likely to complete a STEM bachelor or a (STEM) master. This hints to the idea that this policy change enlarges unequal opportunities for students. Students from wealthier households are more likely to have higher self-esteem (Guyon and Huillery, 2021) which could make them more confident to pursue the *Nature/Tech* field in high school and, consequently, a STEM major. Moreover, students from households with more funds available can more easily compensate the reduction of STEM hours through private tutoring.

A positive side effect of the policy is that far more students graduate high school with a *Nature/Tech* degree. Therefore, these students will have a basic high school level of knowledge on the STEM subjects. This does not seem to influence the longer run effects. There is no evidence in the data that this gain in high school results in different tertiary education choices or, for example, higher income.

Therefore, the long run effect of the policy is that the same number of people are doing STEM majors as before the policy change in the end. These students will have a weaker STEM background due to the less intensive STEM courses in secondary. This would be challenging the effectiveness of this kind of policy which is 'lowering the bar' to a STEM major on a macro level.

On top of that, the composition of the group of STEM graduates has changed with more males and students from wealthier households at the expense of female students with a STEM parent and students from lower income households. This goes both against the goal by *Platform Beta Techniek* to increase the number of (female) STEM graduates and against the idea of equality of opportunity of education.

2 Biased buyers and Market Equilibrium

2.1 Introduction

Consider a market for insurance with two levels of deductibles (out-of-pocket payments): a low deductible (product Low) and the high deductible (product High). When consumers are not homogeneous in terms of default risk, we know that there is adverse selection on insurance markets (Rothschild and Stiglitz, 1976). We expect that consumers who have a higher risk of defaulting are more likely to purchase product Low than consumers who have lower risk, ceteris paribus.

Now, assume that conventional selection of products based on standard preferences is broken by a bias (Masatlioglu and Ok, 2005). Assume there is a non-standard preference for product Low, then some consumers who would prefer product High with standard preferences now prefer product Low. Low-risk buyers purchasing product Low lower the average risk and consequently the costs for product Low. This lowers the price for product Low in a competitive market. What happens to the price of product High depends on the way the bias affects the buyers.

If the bias leads to an homogeneous additional utility value to all buyers, the highest risk product-High buyers purchase product Low. These buyers have lower risk of defaulting than previous product-Low buyers, but higher risk than the remaining product-High buyers. Therefore, the price for product High falls too with this assumption. This means that the average default risk for both products is lower, even when the average risk over all buyers remains the same. This idea is similar to the Simpson (1951)-paradox. In the Simpson (1951)-paradox, the trend of individual groups can be unequal to the trend of the combined group.

If the bias affects only a subset of the buyers at random and we assume that biased buyers always buy product Low, the price of product High stays the same as the average risk of product-High buyers does not change. In that case, the price difference between product Low and product High decreases which leads to some buyers who were not affected by the bias also buying product Low.

The paper introduces a laboratory experiment which simulated an insurance market with buyers and sellers to see how a bias affects the market equilibrium. This market had a mandatory insurance with two deductible levels, product Low and product High. The deductibles were set by the experimenter. In the first phase of a period, multiple sellers set prices for the products simultaneously. In the second stage of the period, buyers bought the insurance from one of the sellers. Buyers in the experiment were heterogeneous in terms of the probability of a damage.

Instead of choosing between product Low and product High directly, buyers had to report their relative valuation (RV) for product Low compared to product High. This is implemented to elicit preferences for product Low. The RV is the premium difference that buyers were willing to pay for product Low (or analogously, the discount that buyers were willing to accept for product High) in order to be indifferent between both products. If the price difference on the market was larger than the buyer's RV, the buyer bought product High. Otherwise, they bought product Low.

Buyers could buy both products from the start in the baseline treatment (control market). Beside the baseline treatment, I introduced two main treatments. Buyers were only exposed to either product Low in treatment Low and to product High in treatment High in the first three rounds. From the fourth round onward, everyone could buy both products. When buyers are initially exposed to only one of the products, the buyers could possibly get a status-quo bias for that product (Samuelson and Zeckhauser, 1988). This would lead to the buyer valuing the product higher than under standard decision making, which I define as 'overvaluation'. Ritov and Baron (1992) separate the status-quo bias into the preference to keep the current state on the one hand and the reluctance of individuals to take action to change this state on the other hand (omission bias). The focus of this experiment lies on the former. In a laboratory experiment, it was shown that when buyers have to choose a deductible, a status-quo bias disappeared after a few periods (Krieger and Felder, 2013). However, in that study the experimenters chose the prices. There is no experimental evidence about the consequences of this bias on a long-run market equilibrium. The questions this paper asks are whether the, in time fading, status-quo bias would affect the prices even after the status-quo bias disappears and whether people to consequently still purchase a different product compared to a market without a bias for any product (i.e. the control market).

The experimental setting is similar to the Dutch health insurance market. Since the Dutch 2005 Health Insurance Act (Zorgverzekeringswet, 2005), every adult in the Netherlands has to buy health insurance from a private insurer. These insurers are not allowed to screen consumers or discriminate in prices. With this insurance comes a mandatory deductible (out-of-pocket payment) of 385 euros in 2023, but insurance companies have to offer an additional deductible of 500 euros (in total 885 euros) in exchange for a discount on the premium. A voluntary deductible of 500 euros yields a discount of 230 euros on the insurance premium on average, as shown in the appendix. Only eleven percent of buyers opted for the voluntary deductible in 2014, while ex post this would have been profitable for 48 percent of the population (van Winssen et al., 2015). This is also a lower bound, as moral hazard would lead to weakly higher health care demand for the 89 percent of the population with a low deductible (Gerfin and Schellhorn, 2006). This institutional setting is described in more detail in the appendix.

One of the reasons why too few people are choosing the high, voluntary, deductible could be that the mandatory deductible is considered to be the default or the reference point (Kahneman and Tversky, 1979). The voluntary deductible is considered the non-standard option. In the Dutch health insurance market, the amount of people choosing the maximum deductible has increased from five percent in 2006 (van Winssen et al., 2016) to eleven percent in 2014 (van Winssen et al., 2015). Between 2017 and 2021, the share of people with a voluntary deductible was between twelve and thirteen percent (van Hijum, 2021). This is still far away from the aforementioned 48 percent and goes against the idea that status-quo bias would fully disappear after a few periods when buyers are accustomed to the new situation like in Krieger and Felder (2013). The question is whether the initial preferences for product Low influenced the current price difference on the market. If the bias led to different market composition and the price difference is smaller than in a bias-free market, then more buyers purchase product Low even if they would have no bias towards product Low.

In my experiment, a first question is whether buyers report a higher RV when endowed with a higher risk. I find that there is significant adverse selection, but it is weaker than ex-ante theoretically predicted. Subsequently, it is observed that an initial exposure to product Low induced a bias for product Low. This means that the buyers overvalued product Low more if they had to buy it in the first three rounds. However, this does not apply to product High. On the contrary, people exposed to product High in treatment High valued product Low consistently higher than buyers in the control market.

The bias affected treatment-Low buyers with the lowest risk of getting a damage more than higher risk buyers. This means that the bias is not additive and the treatment did not only affect buyers close to the indifferent buyer under standard choice, but also buyers with lower risk of getting a damage. This lowers the average risk of a product-Low buyer, but this does not mean that the product-High buyers are also healthier on average as expected in the Simpson (1951)-paradox. If a buyer with the lowest risk buys product Low, this would increase the average risk of the remaining product-High buyers. That buyer would make product Low cheaper and product High more expensive with switching. Product High then yields a lower discount on that market, which makes product High less attractive even for buyers without a bias.

The overvaluation of product Low by treatment-Low buyers is decreasing over time. This is in line with the laboratory experiment of Krieger and Felder (2013). However, we see that sellers have different beliefs about the market composition. Treatment-Low and -High sellers expected the same composition as their counterparts in the control markets in the first round where both products can be purchased. Immediately afterwards, the sellers in the treatment-Low markets expected weaker adverse selection which would lead to lower price differences under perfect competition compared to the control markets.

Another finding of this experiment is that some sellers make small losses on selling product Low. This means that they drive the price difference down even more than what would be expected under perfect price competition. This cannot be explained by standard profit maximization. However, if one seller in an experiment wants to motivate buyers to buy product Low, they can sacrifice a small share of their endowment to make product Low even more attractive.

Taking everything into consideration, despite gradually equalizing RVs across treatments, product Low remains a more attractive option in treatment-Low markets in the longer run. If two identical buyers in the treatment-Low and control market report the same RV in both markets, the difference in price differences could lead to the treatment-Low market buyer purchasing product Low and the control market buyer purchasing product High. This can take over the role of the bias for product Low as this buyer affects the market composition themselves.

The paper continues as follows: Section 2.2 describes the theoretical framework. Section 2.3 explains the experimental design and section 2.4 analyses the results of the experiment. The final section concludes.

2.2 Theoretical framework

In this section, I model an insurance market where heterogeneous consumers have to buy an insurance product with either a low or a high deductible (product L and product H, respectively). First, buyers have standard preferences and in section 2.2.5, a bias is introduced for product L. I show that the indifferent consumer has lower risk with a bias for product L. This means that the average risk for a product-L buyer goes down. Therefore, the average costs and price for product L go down when there is a bias. What happens to the price of product H depends on how the bias affects the buyers. If the bias is additive, the price for product H also goes down. On the other hand, if the bias affects a subset of buyers irrespective of risk level, the price for product H remains unchanged. This model is based on the signaling setup of Rothschild and Stiglitz (1976).

2.2.1 Setup

There is a private market for mandatory insurance with N buyers who have an initial wealth level E and $M > 1$ sellers competing under price competition. Buyer i will incur a damage if $X_i = 1$ with a cost of c. Buyers have a heterogeneous risk of $\theta : P(X_i = 1) = \theta_i$ which is ex-ante privately known to the buyer. Assume there is no moral hazard. This means that there is no way that buyers can increase or decrease their θ through their behavior.

There are two insurance products a buyer can buy: product L and product H. The difference between the two products is the out-of-pocket payment made by the buyer in case of a damage. This induces self selection into the products. The mandatory deductibles are set at $0 \leq d_l < d_h \leq c$. These deductibles are set by a regulator and thus are considered exogenous by the market. Note that a case with a voluntary perfect insurance without any deductible, is covered by the special case $d_l = 0$ and $d_h = c$ as the outside option. In the first stage, sellers simultaneously set two prices $\{p_l, p_h\}$. In the second stage, the buyers choose between the two products to maximize their utility.

2.2.2 Standard consumer choice

This model can easily be solved by backward induction. Consumers have the objective to choose the insurance which maximizes their utility. The sellers have to take this into account when they set prices.

There has to be a degree of risk aversion among consumers to prevent the insurance market from being obsolete. For this framework, consider constant relative risk aversion (CRRA). This is popular in the theoretical literature to assume, but there is also evidence from panel data (Chiappori and Paiella, 2011) as well as laboratory experiments (Levy, 1994) supporting CRRA. The utility for monetary value x will then be according to equation

2.1.

$$u(x, r) = \frac{x^{1-r}}{1-r} \tag{2.1}$$

A consumer with personal risk θ^* and risk aversion parameter r is indifferent between products $L = (p_l, d_l)$ and $H = (p_h, d_h)$ if equation 2.2 holds.

$$\begin{aligned} &U(L, \theta^*, r) = U(H, \theta^*, r) \\ &\theta^* * u(E - p_l - d_l, r) + (1 - \theta^*) * u(E - p_l, r) = \\ &\theta^* * u(E - p_h - d_h, r) + (1 - \theta^*) * u(E - p_h, r) \end{aligned} \tag{2.2}$$

It is shown that $U(L, \theta, r) - U(H, \theta, r)$ is strictly increasing in θ on the relevant domain in the appendix. This means that buyers get relatively more utility from product L if their risk is increasing, ceteris paribus. Consider only the relevant cases where both $p_h < p_l$ and $d_h - d_l > p_l - p_h$ given $d_l < d_h$. In other words, as compensation for the higher out-of-pocket payment in case of a damage when having purchased product H, the consumer with strictly positive θ should get a discount on the insurance premium for product H in order to be indifferent. However, the discount can not be higher than the difference between the deductibles or else product L would never be purchased.

In the relevant cases, $U(L, 0, r) - U(H, 0, r) < 0$ as $p_h < p_l$ and $U(L, 1, r) - U(H, 1, r) > 0$ as $d_h - d_l > p_l - p_h$. This implies that there is exactly one indifferent consumer type θ^* where $U(L, \theta^*, r) - U(H, \theta^*, r) = 0$. The buyers with $\theta < \theta^*$ then prefer product H and $\theta > \theta^*$ prefer product L, which means that there is adverse selection in the market.

2.2.3 Profit maximization problem

Sellers have the objective to maximize profits π_m. However, because the sellers are competing under price competition the total profits per seller cannot be positive in equilibrium. Because of seller symmetry, all sellers will set the same prices for both products. Furthermore, the following two theorems are necessary conditions for the existence of a stable equilibrium.

Theorem 1 *There cannot be a pooling equilibrium where all buyers buy the one of the products.*

The proof is identical to Rothschild and Stiglitz (1976). If all sellers sell only one insurance product to everyone in equilibrium, then this would lead to an opportunity for a seller to offer the other product to persuade part of the consumers to buy that product against a profit. If everyone buys product L, product H can be offered for a slightly lower price. Given that $c - d_h < c - d_l$, the seller has lower costs and the highest-risk buyers stay at product L, which leads to a higher profit. If everyone buys product H, a seller can offer product L for a price high enough to attract the highest-risk buyers (i.e. $\theta^* \approx 1$). Then, the seller cross-subsidizes the losses they make on the marginal product-L market with profits on the large product-H market.

Theorem 2 *In equilibrium, firms can only get zero (expected) profits. On top of that, firms can only get zero (expected) profits for each of the products* $\{L, H\}$.

The formal proof is straightforward and analogous to Rothschild and Stiglitz (1976), Tirole (1988) and Bertrand (1883). For completeness, it is included in the appendix, but the intuition is as follows. Assume without loss of generality that there are two firms (A and B) and firm A has a total profit larger than zero. Firm B can undercut the prices of firm A by $\epsilon > 0$ small enough to gain the whole market and a positive profit. Then firm A would get zero profits, as in standard Bertrand competition, which is a contradiction.

Now assume total profits of firm A are zero, but assume without loss of generality that the profits for product L are larger than zero. Then profits for product H are negative and the firm would be better off not offering product H at all (or offer it at a prohibitively high price). Then firm A would get at least zero profit for product H and there would either be positive profits for firm A or a pooling equilibrium, which are both contradictions.

Therefore in equilibrium, the prices for both products are equal to the average costs and are given in equation 2.3. It will always be that $p_l > p_h$ for two reasons. First, high risk buyers prefer product Low and low risk buyers prefer product High which means that product Low buyers get a damage more often ($E(\theta|L) > E(\theta|H)$). Moreover, the the seller pays a higher share of the damage costs for product Low as $(c - d_l) > (c - d_h)$.

$$p_l = (c - d_l) * E(\theta|L) = (c - d_l) * E(\theta|U(L, \theta, r) > U(H, \theta, r)) \quad (2.3a)$$

$$p_h = (c - d_h) * E(\theta|H) = (c - d_h) * E(\theta|U(L, \theta, r) < U(H, \theta, r)) \quad (2.3b)$$

2.2.4 Example: Buyers with continuous type

In this section, we look at a simple example with a continuum of consumers where $\theta \in [\underline{\theta}, \overline{\theta}]$, with $0 \leq \underline{\theta} < \theta^* < \overline{\theta} \leq 1$. The indifferent consumer θ^* can be found in equation 2.4 and is the solution to equation 2.2.

$$\theta^* = \frac{(E - p_l)^{1-r} - (E - p_h)^{1-r}}{(E - p_h - d_h)^{1-r} - (E - p_l - d_l)^{1-r} + (E - p_l)^{1-r} - (E - p_h)^{1-r}} \tag{2.4}$$

The concept is practically indistinguishable from the Hotelling (1929)-lemma. Every buyer with a θ lower than θ^* has a higher utility for product H. Every buyer with a higher individual risk than θ^* settles with product L as the price difference is not worth the downside risk from the extra loss in case of a damage.

Assume $\theta \sim U[0, 1]$ for illustrative purposes. Then $E(\theta|\theta > \theta^*) = \frac{1}{2} + \frac{\theta^*}{2}$ and $E(\theta|\theta < \theta^*) = \frac{\theta^*}{2}$. As firms would have to charge average costs as they are price competing, equation 2.5 shows the Bertrand-prices for both products.

$$p_l = (c - d_l) * \frac{1}{2} * (\theta^* + 1) \tag{2.5a}$$

$$p_h = (c - d_h) * \frac{1}{2} * \theta^* \tag{2.5b}$$

If equations 2.4 and 2.5 are solved simultaneously given the exogenous parameters $\{E, d_l, d_h, r, c\}$, then this gives the solution for $\{\theta^*, p_l, p_h\}$. For example, for $\{E, d_l, d_h, r, c\} = \{200, 0, 200, 0.8, 200\}$, the equilibrium is $\{\theta^*, p_l, p_h\} = \{0.1591, 115.91, 0\}$. In this example, product L is a perfect insurance (without a deductible) which is bought by around 84 percent of the buyers for a price of 115.91 in equilibrium. Product H is identical to not purchasing insurance as $d_h = c$.

Some sets of exogenous parameters lead to multiple numerical solutions to the equations. If this is the case, then only the solution with the lowest prices can be a stable equilibrium. If a solution exists with higher prices for both products, a seller can lower p_l slightly which reduces θ^* and attracts additional product-L buyers with lower risk. This lowers the expected costs for product L more than the reduction in p_l such that a positive profit emerges. Moreover, as product-H buyers also have a lower average risk level when θ^* falls, the seller also makes a profit from selling product H. Hence, the numerical solution with high prices is not an equilibrium.

It has to be noted that there is no guaranteed existence of a stable equilibrium for the full set of exogenous parameters. In the experiment, a specific discrete buyer distribution is used where an equilibrium exists. The buyer distribution, the exogenous parameters and the predicted equilibrium of the experiment are presented in section 2.3.3.

2.2.5 Biased buyers

Until now, buyers in the framework had standard preferences. In the remainder of the framework, consumers have a bias for product L. The bias leads to a higher (expected) utility for product L than without the bias, or an overvaluation of product L. This bias could be caused by anything, for example a status-quo bias which comes from initial exposure to product L. The implications of two different ways of how the bias can affect the buyers are compared. First, we look at a bias leading to an homogeneous additional utility value to all buyers. This I call an 'additive bias' and leads only to buyers who are relatively close to indifference under standard decision making having a preference for product L. This lowers the average risk level of both product groups, even when the average risk level over all buyers stays the same. Afterwards, a bias is considered which affects only a subset of the buyers at random. If a buyer is affected by the bias, the buyer always purchases product L. I refer to this case as a 'binary bias'.

2.2.6 Case 1: Additive bias

Assume that there is a utility value $\gamma \geq 0$ for product L originating from non-standard preferences, like in Masatlioglu and Ok (2005). As before, we take $\theta \sim U[0, 1]$ as our distribution of the consumers' risk. γ is additive for simplicity as it is both simple and used in the literature (e.g. Altmann et al. (2019)). One can instead assume a proportional, multiplicative term for status-quo bias, but this is not necessary for this illustration.

This γ leads to $U(L, \theta_i, r, \gamma) = U(L, \theta_i, r) + \gamma \forall \theta_i$. This is extra utility because of the bias and is not related to e.g. a 'deductible aversion' as in Pauly (2010). The additive bias gives a new utility function for product L, $U(L, \theta^*, r, \gamma)$, which can be seen in equation 2.6. Utility for product H remains unchanged. Consequently, equation 2.7 shows that the bias term is now in the equation affecting the indifferent consumer.

$$\begin{aligned}
&U(L, \theta^*, r, \gamma) = U(H, \theta^*, r) \\
&\theta^* * U(E - p_l - d_l) + (1 - \theta^*) * U(E - p_l) + \gamma = \\
&\theta^* * U(E - p_h - d_h) + (1 - \theta^*) * U(E - p_h)
\end{aligned} \tag{2.6}$$

$$\theta^*(\gamma) = \frac{(E-p_l)^{1-r} - (E-p_h)^{1-r} + \gamma * (1-r)}{(E-p_h-d_h)^{1-r} - (E-p_l-d_l)^{1-r} + (E-p_l)^{1-r} - (E-p_h)^{1-r}} \tag{2.7}$$

The first order derivative of $\theta^*(\gamma)$ with respect to γ is negative for all $r \neq 1$ and is shown in the appendix. It is important to note that the indifferent consumer θ^* now has a lower risk. Given the prices in equation 2.5 depend on θ^*, the sellers have to adjust their prices as a response. It can be seen in equation 2.8 that the first order derivatives for both prices with respect to θ^* are positive.

$$\frac{\partial p_l}{\partial \theta^*} = \frac{1}{2} * (c - d_l) > 0 \tag{2.8a}$$
$$\frac{\partial p_h}{\partial \theta^*} = \frac{1}{2} * (c - d_h) > 0 \tag{2.8b}$$

This would imply that $p_l(\gamma) \leq p_l(0)$ and $p_h(\gamma) \leq p_h(0)$. Therefore, an additive bias for product L decreases both prices. This can be intuitively explained: the bias induces some consumers around the threshold to switch to product L. These product-L buyers are healthier than original product-L buyers, meaning that the pool of product-L consumers gets healthier on average. Moreover, the pool of product-H buyers lose their relatively sickest consumers and also gets healthier on average. Therefore, like in the Simpson (1951)-paradox, the average risk levels of both products decrease and due to firms getting zero profits in equilibrium both prices fall. The effect on $p_l - p_h$, the discount in insurance premium for having a high deductible, is ambiguous.

The lower insurance premiums do not mean that the expected number of damage claims fall. The expected damage costs for consumer i are $\theta_i * c$ and are not related to deductible choice as there is no moral hazard. This just implies that the people who are affected by the bias are over-insuring compared to standard decision making. Therefore, they pay more for their insurance than under standard decision making and cross-subsidize buyers whose product choices are not affected by the bias.

2.2.7 Case 2: Binary bias

In Case 1, all buyers have an additive and equal bias γ for product L. An alternative assumption would be that not all buyers are affected by a bias, but that the bias only affects a subset of buyers independent of their

θ. This would lead to a different equilibrium.

Now, assume a buyer is biased with probability $\Gamma \in (0,1)$ and always purchases product L when affected. Otherwise, the buyer has no bias and values both products as under standard decision making. This leads to the expected risk levels of both products from equation 2.9.

$$E(\theta|H) = \frac{(1-\Gamma) * E(\theta|\theta < \theta^*)}{1-\Gamma} = E(\theta|\theta < \theta^*) \tag{2.9a}$$

$$E(\theta|L) = \Gamma * E(\theta) + (1-\Gamma) * E(\theta|\theta > \theta^*) < E(\theta|\theta > \theta^*) \tag{2.9b}$$

$E(\theta|H)$ remains unchanged in terms of θ^* and $E(\theta|L)$ is lower as $E(\theta) < E(\theta|\theta > \theta^*)$. As $p_l = (c - d_l) * E(\theta|L)$, the price for product L decreases given θ^*. This would lead to a decreasing $p_l - p_h$, which would lower θ^*. This would then lower the prices of both products a bit more, but at least the price difference is always lower than under standard decision making as that is what drives the decrease of θ^*.

In this example, the price difference between the products is smaller in a market with a bias. This is as not just the marginal consumers are driving down the price for product L as in Case 1, but in this case also some extremely healthy buyers buy product L. Assume there is a buyer who is unaffected by the bias with θ slightly below θ^* under standard decision making. This buyer would purchase product H in a market without bias. However, they would purchase product L if they buy on a market where a fraction of the buyers are affected by the bias. In both cases, the buyer is rationally optimizing their utility, but their product choice is affected by the bias of the other buyers.

2.3 Experimental design

The theoretical model only covers a static environment. It is shown in Krieger and Felder (2013) that a status-quo bias can disappear over time. However, the question that then arises is whether the prices take over the role of the initial bias. Moreover, we can see whether a bias in the insurance market is more likely to be additive or binary. To answer these question, I designed a laboratory experiment. In this section, the experiment is discussed. In section 2.3.1, the design of the insurance market is explained. In section 2.3.2, the treatment differences are presented. In section 2.3.3 the theoretical predictions are made for the parameters of the insurance market. Section 2.3.4 elaborates on the experimental procedures.

2.3.1 General design

The structure of the experiment is as follows. There were twenty participants in a session, who were all assigned a role at random at the beginning of the session. Eight participants were assigned the role of seller and the other twelve participants were buyers. At the beginning of each session, each participant was given a show-up fee of 50 ECUs (with an exchange rate of 10 ECUs = 1 Euro) and the role of either seller or buyer, which they kept throughout the experiment.

The participants were separated in two equal markets (Market 1 and Market 2) of four sellers and six buyers each. The reason to have four sellers in each market is to enforce a competitive environment. Two or three sellers in a market should theoretically also lead to price competition, but collusion might occur in experiments with fewer than four sellers (Dufwenberg and Gneezy, 2000). As sellers who are competing under price competition make zero profits in theory, sellers received a starting capital of 150 ECUs.

At the beginning of the experiment, buyers all got an endowment of 200 ECUs. Moreover, they were assigned a personal risk to get a damage of 200 ECUs with a known probability (θ) between 10 and 50 percent according to the distribution in Table 2.1. Buyers were obligated to buy an insurance on their insurance market from one of the four sellers. This insurance was never perfect and always came with a deductible (d), an out-of-pocket payment made by the buyers in case they suffered a damage. All sellers sold the same two insurance products. One product had a relatively low deductible of 90 ECUs (Product Low, named Product A in the experiment) and one product had a relatively high deductible of 155 ECUs (Product High, named Product B in the experiment). These deductibles were fixed across all sessions.

A session consisted of seven rounds of three periods each. In the first stage of every period, the sellers had to post a price for each of the insurance products $\{p_l, p_h\}$. In the second stage, buyers always had to choose from which seller to buy $\{p\}$ and sometimes also which product they wanted to buy $\{p, d\}$. Which product the buyers could buy and whether they could choose their product depended on the period and the treatment.

When the buyer could choose a product, they were asked to report their 'relative valuation' (RV) for product Low. The RV is the price difference between the products that would leave the buyer indifferent between both products. The RV is asked to elicit buyers' preferences more precisely with a value rather than only with a binary choice. Moreover, the RV is independent from the posted prices of the sellers. The exact wording differed per

Table 2.1: Distribution of buyers risk in all markets

Risk of a damage	Number of buyers
10 percent	1 buyer
20 percent	1 buyer
30 percent	2 buyers
40 percent	1 buyer
50 percent	1 buyer

treatment and can be found in the appendix. If the price difference between the cheapest options for each product in that period was higher than the reported RV, the buyer automatically bought product High. Otherwise, the buyer bought product Low.

Only one of the periods 4 to 21 was paid out per session according to Charness et al. (2016). This period was drawn at random. This was done in order to prevent wealth effects, possible bankruptcy for sellers and buyers trying to hedge their risk. The first round of three periods was considered to be a practice round in which subjects could get used to the insurance market.

Buyer i's pay-off from a period is displayed in equation 2.10 with buyer i suffering a damage with probability θ_i. For seller j, the payoff is equal to their starting capital of 150 ECUs plus their expected profits $E(\pi_j)$ and can be seen in equation 2.11. N_k is the number of buyers that bought product k from seller j, P_k is the price seller j charged for product k, R_k is the average θ that seller j's product k-buyers have and 110 and 45 are the amounts that sellers would have had to pay in case of a damage (the difference between the damage and the deductible, i.e. $200 - 90 = 110$ for product Low and $200 - 155 = 45$ for product High).

The sellers' pay-offs depended on expected profits instead of realized profits, because of the small market size. There can be relatively large differences between expected damages and realized damages with the small market size from the experiment. In reality, firms sell to a large number of consumers, so the ex post number of damages should be relatively close to the ex ante expected number of damages. Therefore, the sellers' payoffs do not depend on whether the buyers actually get a damage to prevent that sellers were exposed to unrealistic risk.

$$\text{Payoff Buyer}_i = \begin{cases} 200 - p_i, & \text{with probability } 1 - \theta_i \\ 200 - p_i - d_i, & \text{with probability } \theta_i \end{cases} \tag{2.10}$$

$$\text{Payoff Seller}_j = 150 + N_{Low} * (P_{Low} - R_{Low} * 110) + N_{High} * (P_{High} - R_{High} * 45) \quad (2.11)$$

Even when sellers receive their expected profits, there remains a realistic possibility that a seller makes a loss. In theory, a loss could be as high as 197.94 ECUs in a single period (by charging 0.01 ECUs for product Low). While it was incredibly unlikely for a seller to incur these grave losses after the practice phase, it was still theoretically possible that sellers made a loss larger than the show-up fee. The starting capital and show-up fee accumulated could fully cover all theoretical losses on the market and prevent a bankruptcy.

2.3.2 Control and treatment sessions

The difference between the treatment and control markets was the availability of the two products for the buyers. In the control sessions, both product Low and product High were available to purchase for all buyers from the first round onward. In the treatment sessions, all buyers could only buy one of the products in the first three rounds. This product was allocated to them at random. Three buyers in each market could exclusively buy product Low (Treatment Low), the other buyers could only buy product High (Treatment High). All four sellers offered the available product to all six buyers within their respective market.

Buyers could choose a different product only in the first period of a round (period 1,4,7 etc.). In all other periods, the buyers could only change sellers but they had to purchase the last product they bought. The reason why buyers could not choose a product every period is to give sellers time to adjust to the constantly changing market compositions by giving them two periods where the market composition remained unchanged. If buyers were free to choose a product every period, it would have been extremely difficult to determine the optimal market price for sellers. In that case, it would have been likely to observe a chaotic price trend as an artifact of the design of the experiment. In real markets, there is much more stability as there are more buyers. This makes this setting with slower mutations of the market composition more realistic for the sellers rather than less realistic.

After the first three rounds, the market was reshuffled in both treatment and control sessions. This was done such that all buyers with prior exposure to product Low ended up in market 1 and those with exposure to product High all ended up in market 2 in the treatment sessions. The two

Table 2.2: Exogenous variables

Variable	Value
Buyer endowment (E)	200
Costs in case of a damage (c)	200
Deductible for product Low (d_l)	90
Deductible for product High (d_h)	155
Number of buyers on a market (N_{buyers})	6
Probability that each buyer gets a damage (θ)	{0.1; 0.2; 0.3; 0.3; 0.4; 0.5}

Table 2.3: Market Equilibrium

Variable	Value
p_l	$110 * 0.4 = 44$
p_h	$45 * 0.2 = 9$
θ^*	0.3
$U(L, 0.3) \approx U(H, 0.3)$	13.077

markets consisted of identical risk distributions both before and after the reshuffling. The only difference between the buyers in the different markets is the product they could buy in the first three rounds.

Two sellers from market 1 always swapped with two sellers from market 2 such that each new market consists of two sellers and three buyers from each old market. None of the sellers knew the product which buyers had previous exposure to. In the control sessions, the markets were reshuffled equally to keep the structure of the treatment and control sessions similar. From round 4 onward, buyers in the treatment sessions had the same options as buyers in the control sessions. This means that buyers could choose between products Low and High by submitting their RVs.

2.3.3 Theoretical predictions and hypotheses

It can be shown that a market equilibrium exists under the assumptions of CRRA utility as in equation 2.1 with $r = 0.8$ (Harrison and Rutström, 2008) and the parameters set as in Table 2.2.

In periods where buyers cannot select a different product, rational payoff maximizing buyers select the lowest price for the product that they are obligated to buy (independent of their θ) and the sellers post prices under price competition based on the average costs for each product. The interesting case is when buyers can choose between both products. With the parameters from Table 2.2, we know that $\theta^* = 0.3$ is the equilibrium with standard rational decision making. Proof that this is the equilibrium can be found in the appendix.

In equilibrium, the average θ of product-Low buyers (in this example 0.4) is higher than that of the product-High buyers (0.2). That means that the buyer RV is dependent on their personal risk. The Adverse Selection Hypothesis would be in line with standard rational decision making.

Hypothesis 1 (Adverse Selection Hypothesis): *Buyers with a higher θ will have a higher RV for product Low.*

The different treatments ensure that buyers only have exposure to one product in the first three rounds of the experiment. This means that buyers in treatment Low are only aware of product Low. Once buyers are able to purchase both products, the question is what their RVs for product Low are. If treatment-Low buyers had a higher RV for product Low, this would mean that there is a bias for the previously endowed product. This is the Bias Hypothesis. On the other hand, if buyers are rational when reporting their RV it should not matter in which treatment market they are.

Hypothesis 2a (Bias Hypothesis): *Treatment-Low buyers have higher RVs for product Low as control buyers and buyers in treatment High.*

For every buyer, one can predict their RV with a CRRA utility function, their θ and their risk preferences r. If a buyer reports an RV that exceeds the RV that one would expect given their θ and r, they are considered to be 'overvaluing' product Low. The buyers' risk levels and treatment market should not affect the overvaluation if there is no bias. If the Bias Hypothesis holds, product Low would be overvalued in treatment Low and the question is whether the overvaluation depends on θ.

Hypothesis 2b (Overvaluation Hypothesis): *The overvaluation of product Low does not depend on θ.*

If the Bias Hypothesis holds and the Overvaluation Hypothesis holds, this would imply that the treatment only affects buyers around the indifference threshold switch products. This leads to a decrease in the average θ for both products and consequently lower the market prices for both products under price competition as in the Simpson (1951)-paradox.

If the Bias Hypothesis holds and treatment-Low buyers with a low θ overvalue product Low more, then some low-risk buyers would also buy product Low. This would then lead to a larger decrease in average costs for product Low compared to average costs for product High.

The experiment takes place over seven rounds. In theory, a buyer should just submit the same RV over time and purchase product Low when the premium for product Low on the market is smaller than their RV and purchase product High otherwise. However, if the Bias Hypothesis is violated, then the question is whether the bias is constant over time or if the buyers weakens over time as in Krieger and Felder (2013).

Hypothesis 2c (Time Hypothesis): *Buyers have the same RV over time across all treatments.*

The Time Hypothesis implies that buyers purchase the same products and the equilibrium is stable over time. When a bias weakens over time, the question is whether the temporary bias was noticeable enough to give sellers different beliefs about the market composition.

Hypothesis 3 (Seller Hypothesis): *The sellers are not affected by the different treatments.*

If the Seller Hypothesis holds, then that would suggest that sellers have the same expectation of buyer behavior and are posting the same prices. This would lead to people reporting the same RV would purchase the same product regardless of the market on which they would purchase it. If the Seller Hypothesis is violated, then two buyers in two different markets who have the same RV could end up buying different products. If this is persistent over multiple periods, then the previous exposure to products can have a long run effect on the market composition and equilibrium.

2.3.4 Experimental procedures

Beside the decisions on the insurance market, more data was collected from the subjects during the experiment. At the beginning of the experiment, subjects were asked to submit their age, gender, nationality, field of study and study year. The second screen was used to collect payment details. Every screen after the second screen had a timer to ensure timely progression.

If the timer expired for a seller, two random prices $p \sim U(0, 200 - d)$ were submitted on behalf of the seller for that period. The seller would lose their starting balance of 150 ECUs if the timer expired. This is a strong incentive for sellers to post prices themselves. If the timer expired for a buyer, the buyer would not buy an insurance that period and they would lose their entire endowment of 200 ECUs, regardless of whether they would have gotten a damage in that period. This strategy is strictly dominated by buying any insurance, even the most expensive one.

During the seller phase, I elicited sellers' beliefs on the expected probability that product-Low buyers and product-High buyers get a damage in their market whenever buyers could report their RVs. This question was implemented to let sellers actively think about the relation between the probability of damage and the product choice of the buyers. Moreover, we can see through the answers whether they actually understood this idea of adverse selection and whether they noticed different buyer compositions in the different treatment markets. The guesses were monetized and the payoff was 10 ECUs per correct guess. A guess was considered correct if it was within a five percentage point margin of the true probability or if no buyer bought the product in that period. One of the guessing tasks' payoffs from the second round onward was paid out at random. This task was drawn independently from the insurance market period which was paid out.

Buyers had to wait to purchase the insurance product until the seller phase was completed. A potential concern is that a bias for the previously endowed product could be weaker if buyers were exclusively focused on purchasing an insurance product. Therefore, buyers were occupied with problems from the second round onward during the seller phase for which they had 60 seconds each. Every problem could earn buyers a payoff and the payoff of one of the problems was paid out at random. This problem was drawn independently from the monetized insurance market period. Also, it was more time efficient to create a profile on the buyers during the seller phase.

From period 4 to period 11, buyers were presented with eight questions from the long adaption of the Frederick (2005) cognitive reflection test (Primi et al., 2015). These questions can be found in the appendix. The payoffs for these problems were 50 ECUs in case a question was answered correctly and 0 ECUs otherwise. From period 12 onward, buyers were given two other problems per period.

The first problem was an element from the multiple price list of the risk aversion task to assess the buyers' risk profiles (Holt and Laury, 2002). The buyer had to choose between a high or a low risk option with a varying probability $p \in [0.1, 1]$ of the good state. The high risk option paid out 110 ECUs in the good state and 6 ECUs in the bad state. The low risk option paid out 60 and 48 ECUs in the good and bad state, respectively.

The payoffs for these problems depended on whether the subject chose the high or low risk option and the state of the world. The full multiple price list can be found in the appendix.

The second problem was an element from the multiple price list of the loss aversion task (Rau, 2015) which is an adaption of Gaechter et al. (2010). In the loss aversion task, the buyer had to choose between accepting or rejecting a bet. If the buyer accepts the bet, they would win 50 ECUs or lose a varying amount $q \in [10, 55]$ ECUs with a fifty-fifty chance. The payoffs for these problems depended on whether the subject accepted or rejected the bet and the state of the world. The full multiple price list can be found in the appendix.

The problems were presented as single entries from these tasks' lists. Also, the entries were presented in a random order to prevent coherent responses to both dilemmas. If a buyer made consistent choices, then the switching point could be estimated as the average between the two pivotal decisions.

Some people made inconsistent choices due to the randomization. In the risk aversion dilemmas 14.6 percent of buyers were inconsistent and in the loss aversion dilemmas 35.4 percent of buyers made at least one inconsistent choice. For these inconsistent buyers, a switching point was estimated with a logistic model following Engel and Kirchkamp (2019). The risk aversion switching point (RASP) was then transformed in the risk aversion parameter r as used above and in Holt and Laury (2002). The switching point from the loss aversion dilemma (LASP) was used to find a loss aversion parameter λ. This was done using another risk aversion parameter α from equation 2.12 as input in equation 2.13 following Rau (2015).

$$\begin{aligned} &.01 * RASP * [\tfrac{1-e^{(-110\alpha)}}{\alpha} - \tfrac{1-e^{(-60\alpha)}}{\alpha}] = \\ &.01 * (100 - RASP) * [\tfrac{1-e^{(-6\alpha)}}{\alpha} - \tfrac{1-e^{(-48\alpha)}}{\alpha}] \end{aligned} \tag{2.12}$$

$$\lambda = \frac{1 - e^{(-50\alpha)}}{1 - e^{(-LASP\alpha)}} \tag{2.13}$$

The experiment has been conducted online with use of z-Tree (Fischbacher, 2007), z-Tree Unleashed (Duch et al., 2020) and oTree (Chen et al., 2016) from 17 March 2021 to 25 March 2022. The participants were recruited with ORSEE (Greiner, 2015) using the subject pool of the University of Mannheim. 160 participants were recruited in 8 sessions. The average earnings were 21.73 Euros and the duration of the experiment was one hour and 45 minutes on average.

As the experiment has been conducted online, there was a virtual video conference-room for the subjects to sign-in, ask questions and report incidents to the experimenter. Because of issues with connecting to the experiment over the internet, three subjects (all sellers) dropped out of the sessions before the end of the practice stage. This lead to an effective sample size of 157 participants. The participants were on average 23.5 years old and 46.5% were female. A more comprehensive summary of the sample can be seen in the appendix.

2.4 Results

In section 2.4.1, buyer behavior is shown. The first question is whether there is adverse selection. If people select products independent from their personal risk, the market composition of the buyers should not change due to the treatment. Afterwards, the treatment effect on the buyers is discussed. If the buyers do not react to the treatments, sellers across different treatments have no reason to change their beliefs about different market compositions or charge different prices.

The behavior of the sellers will be discussed in section 2.4.3. We can compare the actual prices posted by the sellers to the prices that sellers should post if they are perfectly competitive. Moreover, we can look at the predictions to see if sellers noticed different buyer compositions across different treatments. Unless specified otherwise, all tests used in the results section are non-parametric. The Wilcoxon signed-rank test is used for one-sample tests and paired data tests and the Wilcoxon Rank Sum Test is used for two-sample tests.

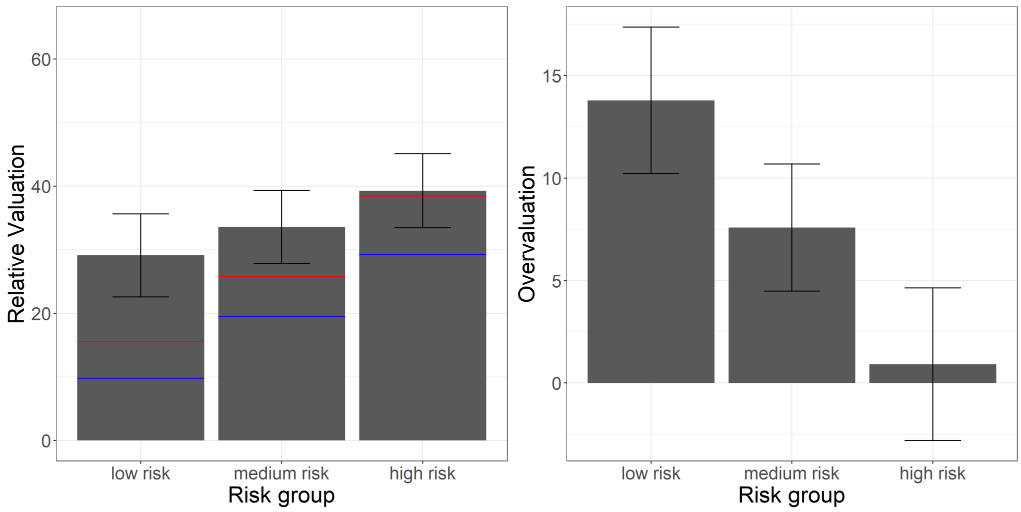
2.4.1 Adverse selection

Do buyers sort into the different products according to their risk profile? There is adverse selection, if buyers with a higher risk of default have a higher RV for product Low. Then buyers are more likely to purchase product Low, ceteris paribus.

The adverse selection can be seen in Figure 2.1 where buyers are split up into three different risk categories. Buyers with $\theta \leq 0.2$ are considered 'low-risk' buyers, buyers with $\theta = 0.3$ are considered 'medium-risk' buyers and buyers with $\theta \geq 0.4$ are considered 'high-risk' buyers. All risk categories consist of one third of the buyers, as there are twice as many observations with $\theta = 0.3$ as with the other risk levels. The adverse selection separated by individual risk levels (θ) can be found in the appendix.

The average RVs for product Low by risk category can be seen in Figure

Figure 2.1: Valuation for product Low



(a) RV for Product Low (b) Overvaluation for product Low

Note: The y-axis in subfigure (a) is the average reported RV for product Low. The y-axis in subfigure (b) is the average overvaluation for product Low. The blue lines are the RV levels for risk neutral buyers and the red lines show the expected RVs corresponding to the groups' average risk-aversion parameter r as measured by the Holt-Laury task. The whiskers indicate 95 percent confidence intervals.

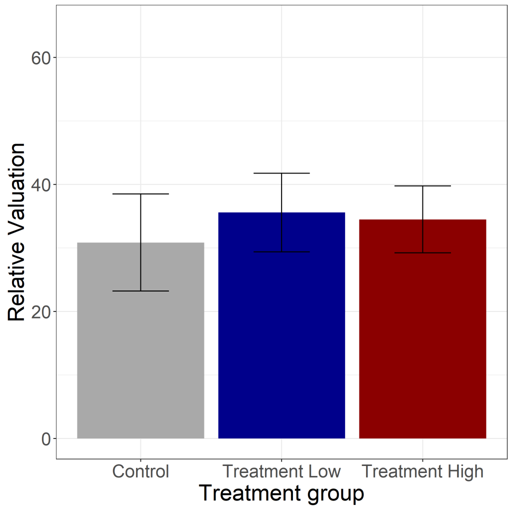
2.1a. Low-risk buyers have a significantly lower RV for product Low (p-value < 0.001) than high-risk buyers. Buyers with medium risk value product Low weakly significantly different than buyers with low or high risk (p-values .083 and .088 respectively).

The theoretical RVs for risk-neutral buyers are are simply $\theta * 65$, where 65 is the difference between the deductibles, and are displayed as blue lines. The average RVs across all risk categories are significantly above these risk-neutral RVs. This is in line with buyers being risk averse on average across all risk categories. The average expected RVs for buyers given their risk aversion parameter are displayed with red lines. Overall, $r = 0.46$, which is lower than expected from the literature.

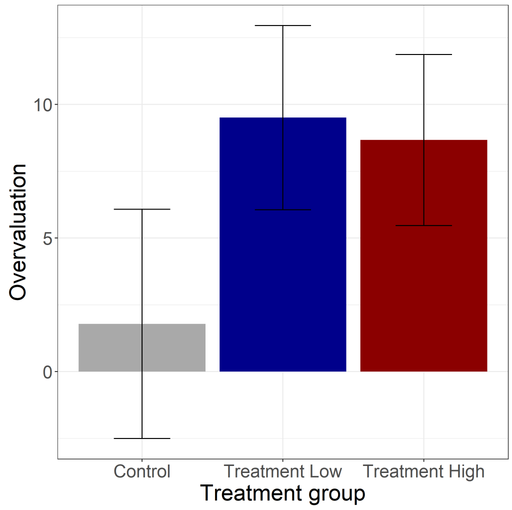
It can be seen that the RV for product Low is higher than what we would have expected given the average risk aversion parameter. Therefore, the average 'overvaluation' for product Low is shown in Figure 2.1b. An individual is considered to be overvaluing product Low, if the RV reported by the buyer exceeds the predicted RV using the CRRA utility function, the individual's risk level and the individual's risk aversion parameter from the risk aversion task in the experiment.

Low- and medium-risk buyers are significantly overvaluing product Low (p-values < 0.001 and 0.007, respectively), whereas this is not the case for

Figure 2.2: Average buyer valuation by treatment



(a) RV for Product Low



(b) Overvaluation for product Low

Note: The y-axis in subfigure (a) is the average reported RV for product Low. The y-axis in subfigure (b) displays the average difference (in ECUs) between an individuals reported RV for product Low and the RV a buyer should have with their risk aversion parameter from the risk aversion task using a CRRA model. The whiskers indicate 95 percent confidence intervals.

high-risk buyers (p-value 0.963). Moreover, low-risk buyers significantly overvalue product Low compared to high-risk buyers (p-value = 0.009). This implies that the adverse selection is significantly weaker than expected due to low-risk buyers overvaluing product Low.

Observation 1: *The RVs between the risk categories differ significantly, which implies that there is adverse selection. However, the difference is less than expected due to the overvaluation of product Low by low- and medium-risk buyers.*

2.4.2 Treatment effect on the buyers

Do the different treatments significantly impact the RVs of the buyers? As the treated buyers have prior exposure to one of the two insurance products for the first three rounds of the session, they could have developed a preference for their endowed product.

The average (excess) RVs from round 4 onward are displayed in Figure 2.2. The RVs of treatment-Low and treatment-High buyers are not significantly higher than of the control buyers (p-value .181 and .242, respectively). However, buyers in treatment Low and treatment High do statistically overvalue product Low (p-values .006 and .004, respectively). The average overvaluation for product Low by treatment is shown in Figure 2.2b. This implies that both treatments induced buyers to overvalue product Low. Unexpectedly, buyers in treatment High overvalue product Low despite being exposed to product High initially. This is not compatible with the status-quo bias hypothesis.

Observation 2a: *Buyers in treatment Low and treatment High significantly overvalue product Low.*

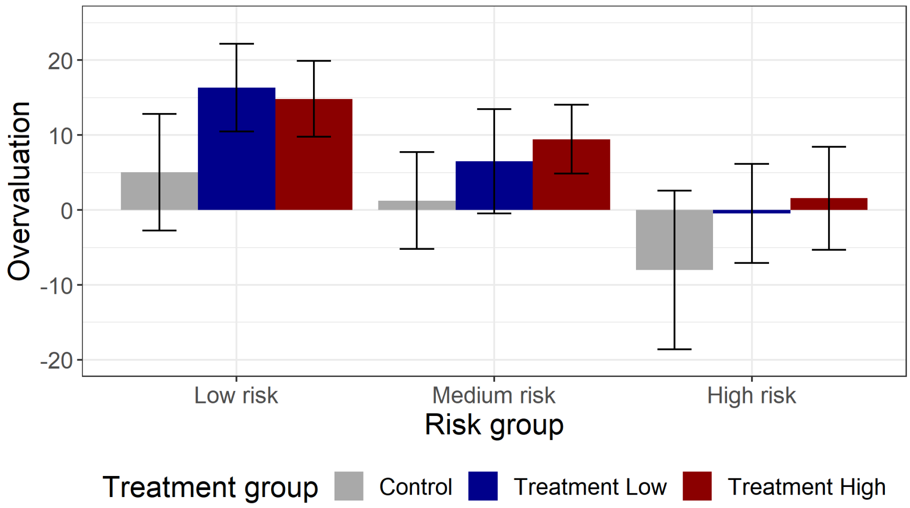
Is there an interaction between the overvaluations of product Low of the low- and medium-risk buyers and the treatment buyers? In other words, would low-risk buyers in treatment Low have a higher overvaluation? If there is no interaction, then the treatment affected all buyers similarly and the overvaluation of product Low could simply be due to probability-dependent risk preferences (Fehr-Duda and Epper, 2012). If there is an interaction, then the treatment did not lead to an additive or multiplicative bias, but would affect low-risk buyers even more.

The overvaluation separated by risk group and treatment can be seen in Figure 2.3. Low-risk buyers in treatment Low and treatment High do significantly overvalue product Low (p-values 0.016 and 0.012, respectively), while control buyers with low risk do not overvalue product Low significantly. Medium-risk buyers overvalue product Low to a lesser extent in the treatment markets (p-values 0.082 and 0.012, respectively). High-risk buyers do not seem to overvalue product Low in either treatment (p-values 0.569 and 0.791, respectively).

This means that the overvaluation from low-risk buyers in Figure 2.1b comes mainly from the treated buyers and is caused by the different treatments and not because of probability dependent risk preferences (Fehr-Duda and Epper, 2012). If low-risk control buyers also significantly overvalued product Low, buyers would simply over-weigh the relatively small probability of getting a damage, but now only the treatments made product Low more attractive for low-risk buyers.

Observation 2b: *The treated buyers with low risk and medium risk significantly overvalue product Low, while buyers with high risk do not overvalue product Low.*

Figure 2.3: Overvaluation for product Low



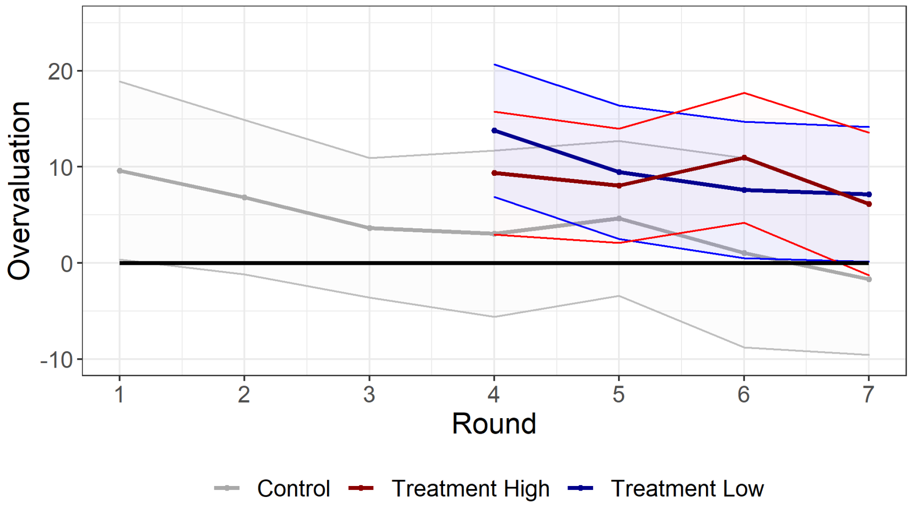
Note: The y-axis shows the average difference (in ECUs) between an individuals reported RV for product Low and the RV a buyer should have with their risk aversion parameter from the risk aversion task using a CRRA model split by treatment and riskgroup. The whiskers indicate 95 percent confidence intervals.

Low-risk buyers who have previously been exposed to product Low significantly overvalue product Low when they have to choose between the two products. This could be a bias for the previously endowed product. However, this can not explain the overvaluation for product Low by treatment-High buyers. If they had a status-quo bias for product High, they would have undervalued product Low. Therefore, the treatment-High buyers either do not have a preference for their status-quo, or any status-quo bias for product High got overshadowed by another bias that led buyers to the overvaluation of product Low.

So far, we only considered average buyer behavior. As it is known that a status-quo bias can disappear after a few periods (Krieger and Felder, 2013), it is interesting to analyze the difference between the RVs between round 4 and round 7. The question is whether the overvaluation remains constant over time or if it is higher in round 4 than in round 7, as one would suspect with a status-quo bias.

In round 4, the first decision was made by treated buyers after they had been exposed to their initial product during first three rounds. In round 7, buyers have had exposure to both products for some time and might even have purchased the other product at some point. This could have an impact on the buyers' status-quo bias caused by the treatment. The RVs over time by treatment are shown in Figure 2.4. The general movement over time is

Figure 2.4: Overvaluation for product Low over time by treatment



Note: The y-axis shows the average difference (in ECUs) between an individuals reported RV for product Low and the RV a buyer should have with their risk aversion parameter from the risk aversion task using a CRRA model split by treatment group. The shaded areas indicate 95 percent confidence intervals.

that the RVs in round 4 are significantly higher than in round 7 (p-value = .005).

The RVs generally fall in the later round. However, there are differences between the different treatments. The RVs drop across all treatments. However, the decrease is not significant in the control markets (p-value .163) and treatment-High markets (p-value .353). Treatment-Low buyers do value product Low significantly lower in the last round (p-value .009). The overvaluation also significantly decreased over time, as the overvaluation within subject is just a level shift. This is in line with status-quo bias disappearing over time for treatment-Low buyers. Moreover, as status-quo bias is not happening to treatment-High buyers, it makes sense that their RV does not decrease.

Observation 2c: *The average relative valuation for product Low of treatment-Low buyers is decreasing over time.*

To separate the treatment effects from confounds, we can use regressions from equation 2.14. The multiple regressions are used to look at the treatment effect on the RV or overvaluation of buyer i in round t while controlling for the age, gender, nationality, study major, CRT score, risk aversion parameter (r) and loss aversion parameter (α). Note that in the regressions where overvaluation is the dependent variable, the risk aversion parameter is omitted from the set of controls as r is part of the construction of the dependent variable.

$$RV_{i,t} = \alpha_0 + \alpha_1 * T_i + \alpha_2 * \theta_i + \gamma_1 * X_i + \epsilon_i \tag{2.14a}$$

$$Overvaluation_{i,t} = \beta_0 + \beta_1 * T_i + \beta_2 * \theta_i + \gamma_2 * X_i + \upsilon_i \tag{2.14b}$$

The treatment effect on the RVs can be seen on the left-hand side of Table 2.4. The RVs for treatment Low are significantly higher than for the control group in round 4. This difference is no longer significant in the last round. Treatment-High buyers never had a significantly higher RV compared to the control group. Moreover, the risk level has a significant effect on the RV, which again implies that there is adverse selection.

The next thing to consider is the overvaluation of product Low on the right-hand side of Table 2.4. The treatment-Low buyers significantly overvalue product Low compared to the control group in round 4. This difference is still significant on a ten-percent level in the last round. Treatment-High buyers never significantly overvalued product Low compared to the control group. Moreover, the risk level has a significant negative effect on the overvaluation. This means that the overvaluing is indeed more prevalent with low-risk buyers.

2.4.3 Seller responses

We know now that treatment-Low buyers have a temporary higher overvaluation of product Low. Do the sellers react to the different behavior of the buyers? Seller responses are first analyzed independent from their treatment market. As sellers are not treated directly, it is not necessary to split by treatment when controlling for behavior of the buyers. Afterwards, we look at different expectations from sellers based on different treatment markets.

Sellers predict the market composition in the first period of every round and they set prices in every period. A round consists of three pricing periods. There are two moments of particular interest to consider in a round. In the first period, sellers have to estimate the probability that a product-Low and product-High buyer get a damage. In the last period, sellers are aware of the buyer compositions which means that prices show how competitive sellers are.

The average predictions over time can be found in Figure 2.5a. Sellers

Table 2.4: Treatment Effect on (over-)valuation of product Low

	Dependent variable:					
	RV			Overvaluation		
	(Average)	(round 4)	(round 7)	(Average)	(round 4)	(round 7)
Treatment High	5.335	5.379	5.850	6.838	6.257	7.448
	(4.685)	(5.213)	(5.321)	(4.921)	(5.324)	(5.421)
Treatment Low	7.262	10.790**	7.716	8.260*	11.292**	8.711*
	(4.484)	(5.089)	(5.044)	(4.835)	(5.278)	(5.159)
Risk level (θ)	.335***	.357***	.291*	-.373***	-.371**	-.421**
	(.118)	(.137)	(.152)	(.143)	(.154)	(.165)
Constant	24.896*	27.578*	18.432	22.906	26.796*	15.830
	(13.510)	(14.752)	(16.200)	(14.304)	(15.206)	(16.812)
Observations	88	88	86	88	88	86
Control variables	YES	YES	YES	YES	YES	YES
Adjusted R^2	.149	.161	.042	.121	.130	.062

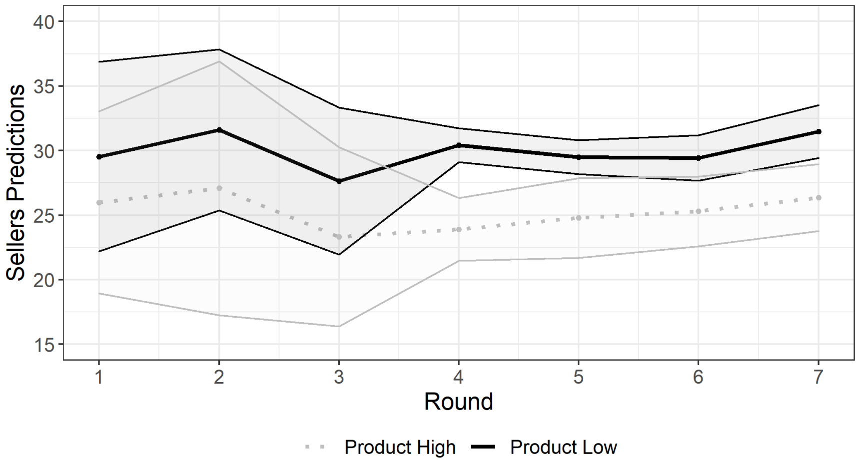
Note: *p<0.1; **p<0.05; ***p<0.01

already predicted in their first prediction that product-Low buyers had a higher chance of getting a damage than product-High buyers (p-value .006). This means that sellers predicted adverse selection from the start of the experiment onward. In the last prediction, the difference is still significant (p-value .007).

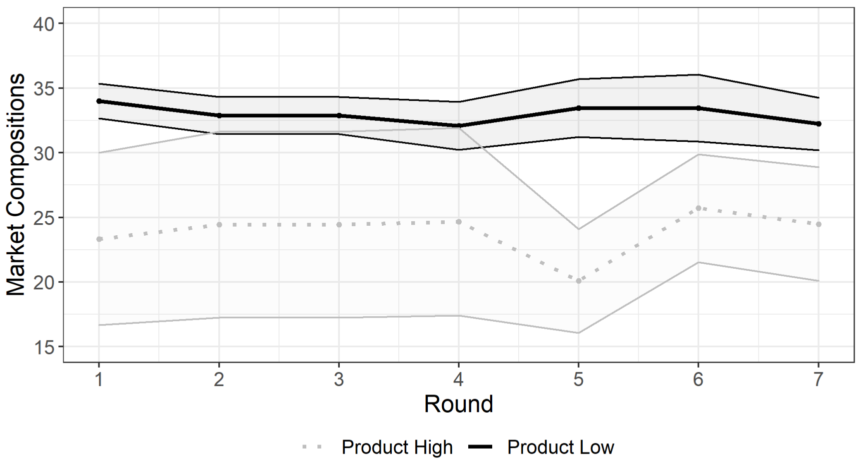
The actual market composition can be seen in Figure 2.5b. We have seen a significant increase in RVs with respect to θ as predicted theoretically. Even when the adverse selection is weaker than theoretically predicted, the adverse selection in buyer composition is present from the first round where buyers report RVs (p-value .006) until the last round (p-value .025). The standard errors for product High are substantially larger as only 27 percent of the time product High is purchased across all rounds and treatments when buyers could report their RVs.

Sellers are fairly accurate with their predictions on average. However, sellers predict a bit less adverse selection than in reality. Sellers underestimate the probability that a product Low buyer gets a damage by 7.6 percent (2.5 percentage points) and overestimate the probability that a product High buyer gets a damage by 7.3 percent (1.7 percentage points). Sellers received potential earnings if their guess was within five percentage

Figure 2.5: Predictions and market compositions



(a) Seller predicitions



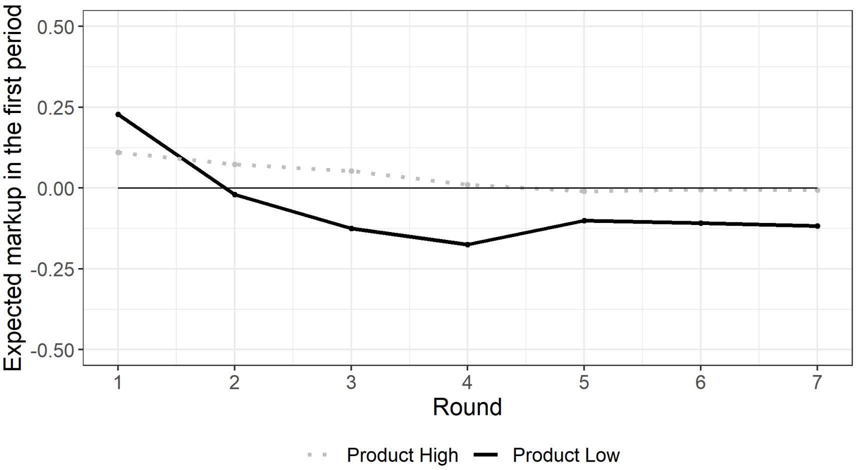
(b) Market compositions

The y-axis of (a) shows the average of the submitted seller predictions on market level separated by product. The predictions are regarding the average buyer risk for the product in their market as a percentage value. The y-axis of (b) shows the average of the buyers theta on market level as a percentage value separated by product. The shaded areas indicate 95 percent confidence intervals.

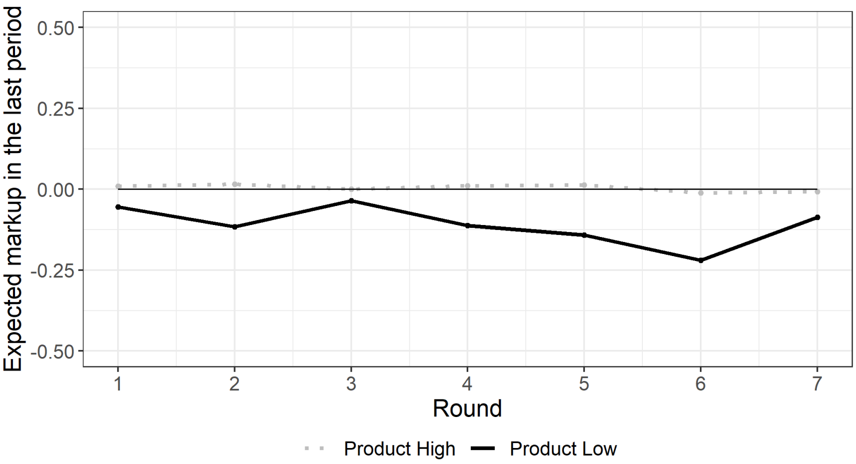
points of the actual average risk. On average, 66 percent of predictions about product-Low buyers and 51 percent of predictions about product-High buyers were within the range. As the number of product-High buyers was relatively low, it was harder to correctly predict the average risk of product-High buyers.

Now, the question is whether the sellers matched their prices with their predictions and the market compositions. Sellers could use their prediction in the first period and the composition in the later periods to compute

Figure 2.6: Expected markups



(a) Expected markup in first period



(b) Expected markup in last period

The y-axis of (a) shows the expected difference between the posted prices in the first period of a round and the expected costs given the submitted seller predictions separated by product The y-axis of (b) shows the expected difference between the posted prices in the last period of a round and the expected costs given the submitted seller predictions separated by product.

the expected costs to sell insurance and set their price accordingly. The percentage difference between the posted price and the expected costs is the expected markup for sellers. If the markup is positive, a sale would lead to a profit for a seller. Analogously, if they sold a product with a negative markup, they made a loss. If sellers competed under price competition, the markup should be zero.

What markups should we compare? It is not ideal to look at a median markup on the market as prices with low markups are more likely to be chosen than prices with high markups. In the experiment, buyers bought insurance from the seller with the lowest price 89 percent of the time. As markups are normalized across markets, we can look at the expected markup across all markets. This is the average of all markups weighted by the probability that this markup is the lowest out of four randomly drawn prices without replacement.

The expected markups for sellers are shown in Figure 2.6. The results for the first period of a round can be seen in Figure 2.6a. In the first period of a round, I compare posted prices to the costs predicted by sellers themselves. For product High, the expected markup was positive and gradually moved toward zero over time. On the other hand, the average expected markup for product Low is significantly lower than the markup for product High (p-value .009) and negative with the exception of the practice round 1.

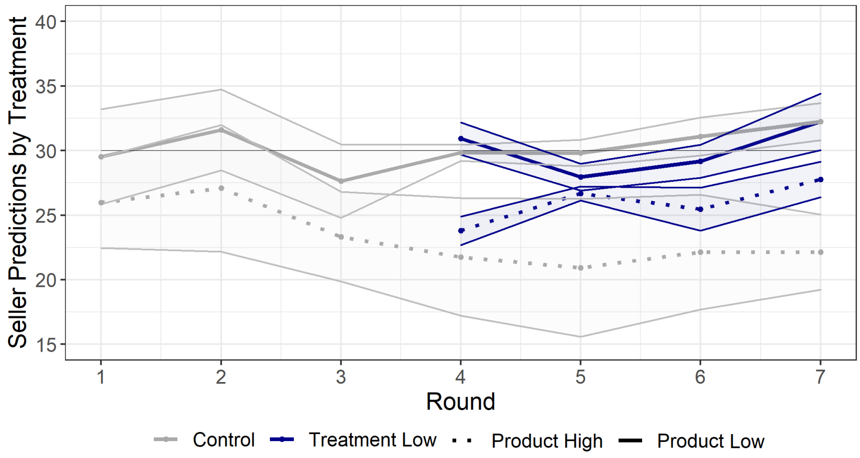
In round 4, the first round where sellers in all markets are predicting market compositions, the expected loss for selling product Low is the largest with 17.4 percent. As the first period is crucial for determining the market composition in that round, the expected losses from product Low led to a lower price difference between the two products and consequently led to more buyers buying product Low.

The expected markups in the last period of a round are displayed in Figure 2.6b. In the last period, the market composition was known to the sellers. Therefore, I compare the posted prices to the actual average costs for insuring a buyer on the seller's market. The average expected markups are still negative for product Low and approximately zero for product High. Also in the final period of a round, the difference between the two products is significant (p-value $<.001$).

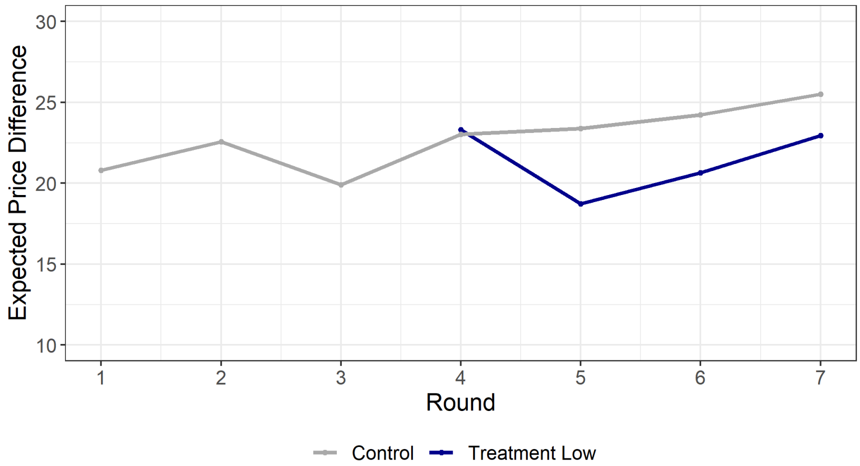
After the composition of the market is known to sellers, any expected profits for selling product High in the early rounds disappeared. Moreover, some sellers continuously overcompeted with undercutting prices for product Low and were willing to accept a small loss to get a high market share even when the buyer compositions were known. However, the buyer compositions were not changed because of underpricing in this period. Underpricing in the last period of a round simply caused a one-to-one cross-subsidy from the seller to their buyers.

The last interesting question is whether the sellers noticed the treatment differences. For this, we can directly look at the sellers predictions. As treatment-High buyers overvalued product Low to a similar extent as treatment-Low buyers, sellers could not have been able to differentiate between buyers from the different treatment markets. Therefore, only the

Figure 2.7: Treatment differences



(a) Seller predictions separated by product and treatment



(b) Expected price differences from seller predictions separated by treatment

The y-axis of (a) shows the average of the submitted seller predictions on market level separated by product and treatment. The predictions are regarding the average buyer risk for the product in their market as a percentage value. The shaded areas indicate 95 percent level confidence intervals. The y-axis of (b) shows the difference in expected costs given the submitted seller predictions separated between products by treatment.

treatment-Low and control markets are compared to each other. The results are similar when all treatment-buyers are compared with control buyers.

The predictions separated by treatment are seen in Figure 2.7a. It makes sense that the predictions are not significantly different between the treatment-Low and control sellers in round 4 (p-value .714), as sellers would not have had a reason to believe that buyers would react differently. Treatment-Low sellers predict a significantly higher risk than control sellers for product-High buyers in round 7 (p-value .017). This is as treatment-Low sellers predict the average product-High buyer to have increasing risk levels over time (p-value .012). Sellers do not predict different buyer compositions for product Low across different treatment markets or over time.

The increased average risk of product-High buyers is in contrast to the prediction with a binary bias in the theoretical framework. In the theoretical framework, the binary bias leads to a decrease in the average product-Low buyer's risk level while the average risk of the product-High buyer remains unchanged. This contradiction can be explained by product High being less popular (27 percent of all insurances sold) than predicted (50 percent), among others due to the expected discounts for product Low given by sellers. This means that in an average market with six buyers, only two buyers bought product High. If the product-High buyer with the lowest risk bought product Low due to the bias, the average product-High buyer's risk increased substantially to the risk level of the highest-risk product-High buyer. If there would have been a continuum of buyers, the theoretical prediction would be more accurate as the affected buyers' risk levels would then be more in line with the average product-High buyer's risk level. However, in both the theoretical case and the results from the experiment, the bias led to less adverse selection. This is due to the bias not only affecting marginal buyers, but also low-risk buyers.

Treatment-Low sellers predict weaker adverse selection than control sellers. Consequently, the expected price difference between product Low and High is lower after the sellers notice the treatment effect, which can be seen in Figure 2.7b. The expected price difference in treatment Low dropped by approximately 20 percent from 23 to 18 ECUs in round 5. This means that a buyer with a constant RV of 20 ECUs would buy product High in both markets in round 4, but they would buy product Low in a treatment-Low market in round 5. Moreover, this buyer would then also influence the market compositions in that round and consequently the sellers' predictions for the subsequent round.

Observation 3: *Treatment-Low sellers predict over time that product-High buyers have higher average θ, which leads to lower discounts under perfect competition.*

2.5 Discussion and Concluding remarks

In the Dutch health insurance market, there is an institutionalized default deductible. Moreover, a radical change in 2006 made the high deductible a new high risk option from the reference point. We should then expect a status-quo bias according to the psychology and behavioral economics literature. If this is indeed the case, what would be the consequences of this status-quo bias on the market equilibrium over time?

I ran an experiment to investigate the dynamic effects of introducing a default product on the market equilibrium. The experiment simulates an insurance market with both buyers and sellers. I elicit the buyers' preferences for the product they were initially endowed with. If buyers' preferences are higher with than without default, this would be in line with status-quo bias. Moreover, the experiment shows whether this status-quo bias is constant or decreasing over a few rounds. The last question is whether sellers observe the (initial) status-quo bias from buyers by changing their expectations about the composition of the market.

When buyers are able to buy both products, buyers who could previously exclusively purchase product Low are willing to pay significantly more for product Low than buyers who could buy both products from the start. This is in line with a status-quo bias. Buyers previously exclusively exposed to product High also find product Low more valuable when they have the opportunity to purchase this product. This contradicts the hypothesis for a status-quo bias for treatment High buyers.

The overvaluation for product Low by treatment-Low buyers decreased over time, which means that the status-quo bias weakened. This led to more buyers purchasing product High. However, a temporary status-quo bias still has an influence on the long run equilibrium. As buyers with the lowest risk to get a damage are overvaluing product Low the most, the low-risk treatment-Low buyers buy product Low.

As a result, sellers predicted weaker adverse selection in the treatment-Low markets. This leads in equilibrium to lower price differences than in the markets without a bias for a product. Even a buyer who is not exposed to the treatment, would buy product Low instead of product High in a market with a lower price difference and influences the market composition. This effect is amplified by sellers who sold product Low at a loss-making price which incentivized even more buyers to buy product Low.

It is unclear how a regulator should respond to an equilibrium that is affected by a bias. The bias would imply that certain individuals are over-insured on the insurance market. This would mean that these influenced buyers are cross-subsidizing the buyers whose product choices are not affected by the bias. The fact that only product Low becomes cheaper, implies that the discount decreases and this affects the equilibrium through weaker adverse selection.

The lower price difference can take over the role of the bias even when the bias would no longer affect the buyers' valuations. Moreover, in a model with moral hazard, which is outside the scope of this project, over-insured individuals could make higher societal costs as the costs of making an insurance claim are lower when a consumer has a lower deductible. This may go against one of the motives of a regulator to introduce a mandatory deductible in the first place.

3 Face Masks and Economic Behavior

3.1 Introduction

During the COVID-19 pandemic, the majority of countries imposed (temporary) mask mandates on the street, in public spaces or at work (CDC (2020); ECDC (2020); RKI (2020)). Yet, initially there was no global political consensus on face mask mandates (RIVM (2020); WHO (2020)). This was mainly due to heterogeneous evidence of the effectiveness of face masks in reducing infections (Xiao et al. (2020); Leung et al. (2020)). During the pandemic, evidence was provided that the face mask mandates did indeed lower COVID-19 transmissions (Lyu and Wehby (2020); Mitze et al. (2020)). However, the arguments in the discussion on mask mandates were considering only the medical effectiveness of the face masks.

The discussion did not take into account if wearing a mask has consequences on economic behavior. One might not immediately see why masks could affect economic behavior. However, it has already been shown that wearing masks can lead to altered breathing patterns (Louhevaara, 1984) and more self-reported discomfort in fatigue, hotness, tightness and breath resistance (Fikenzer et al., 2020) during physical labor. Previous studies in the economic literature have already established that fatigue (e.g. Viner et al. (2008), Abd-Elfattah et al. (2015)), air pollution (Archsmith et al., 2018), hotness (Heyes and Saberian, 2019) and lack of fresh air (Chen and Schwartz, 2009) have an influence on economic behavior.

This paper uses a laboratory experiment to investigate whether disposable face masks affect economic decision making, social preferences and productivity of healthy young adults sitting behind a computer. One paper finds that cognition of children is not affected by wearing masks in a classroom (Schlegtendal et al., 2021). However, in this paper I look at whether people become less productive and if their individual and social preferences and economic characteristics are affected when they have to wear a face mask.

Subjects who wore a mask were 5.6 to 7.1 percent less productive in a short task with high intensity. Moreover, this effect is larger if the task is done later in the experiment, when subjects were wearing their masks longer. On the other hand, there is no loss in productivity when a mask is worn during a productivity task with low intensity.

With regard to economic interactions, I find that proposers in an ultimatum game increased their offer by close to twenty percent if they had been asked to wear a mask. However, as no other measures changed, it is unlikely that the offers are improved because of generosity. It is plausible that the offers are higher due to different beliefs on acceptance of the offer. In that case, it is possible that the proposers would have anticipated the idea that an offer accompanied with a positive facial expression, is more likely to be accepted (Mussel et al., 2013). However, there is no evidence that an offer from a mask wearer is less likely to be accepted. Cognition, risk and loss attitudes, honesty and generosity are not affected by mask wearing.

Beside the role masks play in the debates around ways to combat a viral pandemic, this research contributes more generally to the 'biology and economics' literature. In this field, economic agents are modeled as 'wet machines' who are susceptible to their surrounding environments. Research in this area has shown among others that hunger can reduce mental function under adults (Weaver and Hadley, 2009) and children (Weinreb et al., 2002) and under animals it can influence perceptions of risk (Ferrarelli, 2016). Also poverty can reduce cognitive function and reduce decision quality (Mani et al., 2013). No research has been done looking into the effect of face masks on economic behavior.

The experiment consisted of three parts: a part on individual characteristics, a productivity part and a part on interactions. The part on individual characteristics assesses cognition (Raven, 2003), risk aversion (Holt and Laury, 2002), loss aversion (Rau, 2015), honesty (Crumpler and Grossman, 2008) and generosity (Fischbacher and Foellmi-Heusi, 2013). The productivity part required subjects to count zeros for 14 miunutes (Abeler et al., 2011). The interactions part includes a competition game (Niederle and Versterlund, 2007) with a two-minute slider task (Gill and Prowse, 2012), a trust game (Berg et al., 1995), an ultimatum game (Güth et al., 1982) and a standard public goods game (Ledyard, 1995).

The order of the three parts is reversed at random sessions to allow for analyzing the differences in wearing a mask at the beginning of a session compared to wearing one at the end of a session. This is done to answer the question whether wearing a face mask for a shorter or longer consecutive period enhances any effect on social preferences, individual characteristics or productivity.

This paper is organized as follows. Section 3.2 briefly describes the literature existing on face masks and the 'biology and economics' agenda. Section 3.3 explains the design of the experiment and section 3.4 shows the results of the experiment. The final section concludes this study.

3.2 Literature

The question this paper asks is whether the effect of wearing face masks on channels as fatigue and air conditions is sizeable enough to directly affect economic behavior of individuals who are sitting inside behind a desk. In the literature of other fields than economics, the effect of face masks are usually in terms of exercising or physical labor. As this research does not focus on the mechanisms that could drive the effects of the face masks, it is worth discussing the research on the physiological effects of face masks. Several studies have researched the effect of face masks on physiological factors. In these studies, the masks are worn during exercise or physical labor.

There is already a trade-off illustrated between protecting workers' health and the performance of workers when respirators are used (Johnson (2016); Louhevaara (1984)). The direct effects of the respirators are alterations in breathing pattern, retention of carbon dioxide and hypo-ventilation (Louhevaara, 1984). The consequences are that maximal physical work capacity would be reduced.

One of the channels through which masks could influence economic choices is tiredness. Tiredness reduces self-control (Kahol et al., 2008), it increases risk-taking (Viner et al., 2008) and according to Tchen et al. (2003) and Abd-Elfattah et al. (2015)) it also reduces cognitive function. In the physiological literature, it has been found that wearing a disposable mask leads to more self-reported discomfort in fatigue, hotness, tightness and breath resistance (Fikenzer et al., 2020). Moreover, physical work while wearing protective facial gear causes increased skin temperature and also more self-reported discomfort (White et al., 1991).

In this experiment, the differences between laboratory sessions with or without face masks are analyzed. All tasks were done while participants were sitting behind a computer instead of doing physical exercise. Wearing a mask during the experiment is related to being exposed to different air conditions. A lack of fresh air has already been linked to reduced cognitive function (Chen and Schwartz, 2009) and mood (Cunningham, 1979). Heyes and Saberian (2019) find that judges are influenced by the outside weather when making decisions on immigration applications, even in air-conditioned rooms. Besides outside temperature, air pollution can have influences on high skilled decision making by for example referees (Archsmith et al., 2018).

In economic literature, it has already been established that behavior can be influenced by aforementioned channels through other means than wearing a face mask. In the literature, decision making is influenced by unwanted changes in mood (Englich and Soder, 2009), cognition (Dijksterhuis et al. (1996) ; Wyer and Carlston (1979)), emotion (Simon, 2012) and fatigue (Pijpers et al., 2007).

3.3 Experimental design

The goal of the experiment is to uncover if there are any differences in choices and performance in various well-known tasks when subjects are wearing a face mask or not and whether the effects of face masks are similar when wearing them continuously for a shorter or longer period of time. The experiment commenced with general instructions, which can be found in the Appendix, and it consisted of ten tasks from the literature. These ten tasks were separated in three parts: an individual characteristics part, a productivity part and a social interactions part. For every task, subjects earned ECUs (experimental currency units), with 8 ECUs being worth 1 Euro.

Only the earnings of one out of the ten tasks got paid out to the subjects on top of their show-up fee of 5 Euros. This task was selected separately for each subject by the computer completely at random. The advantages of paying out only one task are twofold (Charness et al., 2016). Firstly, there were no income effects or incentives for people to hedge across tasks. Secondly, every task could get higher potential stakes compared to paying out all tasks, even when the expected payment stayed the same.

After the ten tasks, subjects were asked for their age, gender, nationality and field of study, and they answered some questions about the experiment and their feelings on a five-point Likert scale in a survey. After filling out the survey, subjects received the information on the selected task and their payoff of the task in ECUs. Subjects got paid in cash and in private directly after the experiment. The experimenter did not receive information on the selected task and subjects were made aware of this.

The experiment has been conducted at the University of Mannheim (mLab) from 23 September 2021 until 11 October 2022. This experiment was programmed in oTree (Chen et al., 2016). In total, 160 participants were recruited from ORSEE (Greiner, 2015) who participated in 14 sessions. Every session consisted of a multiple of four participants, with at least eight participants in every session. This was a requirement as two of the social interaction tasks needed groups of four and two other tasks were done in couples. The average duration of a session was 75 minutes and the average subject earnings were 17 Euros.

3.3.1 Part A: Individual characteristics part

This part contains five tasks which create a profile on cognition, risk- and loss aversion, generosity and honesty. Cognitive ability is measured with a selection from Raven's progressive matrices (Raven, 2003). Risk aversion is measured with a standard multiple price list from Holt and Laury (2002) and to assess loss aversion, a list with fifty-fifty bets from Gaechter et al. (2010) and Rau (2015) is used. Generosity is measured with a donation decision to a local children's hospice (Crumpler and Grossman, 2008). Honesty is measured by self reporting the outcome of a die roll (Fischbacher and Foellmi-Heusi, 2013).

The original Raven's progressive matrices test (Raven, 2003) consisted of five sets of twelve 3-by-3 matrices containing cells with geometric images. The last cell of a matrix was missing and the goal for the subjects was to find the missing cell out of six to eight options. In this experiment, only the first, second, fifth, sixth, seventh, eighth, eleventh and twelfth matrix of sets C, D and E were used. The subjects had 4 minutes per set and earned 6 ECUs per correctly solved matrix.

The risk aversion test (Holt and Laury, 2002) consisted of a simple multiple price list with 10 choices. Subjects had to choose between a relatively risky option A or relatively safe option B. When option A was chosen, with probability $x \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ a subject earned 230 ECUs and 6 ECUs was earned with probability $1 - x$. When option B was selected, with probability x a subject earned 120 ECUs and 96 ECUs was earned with probability $1 - x$. The subjects had 3 minutes to make all 10 choices. If a subject did not make a choice when the timer expired, they earned 0 ECUs for that choice. One out of the 10 choices got drawn by the computer completely at random to be payoff relevant for this task. This task's earnings were then determined according to the subject's choice and the state of the world.

The loss aversion test (Rau, 2015) was an adaptation from Gaechter et al. (2010) and is similar to the multiple price list of Holt and Laury (2002). This test also consisted of a simple list of 10 choices. At the beginning of the task subjects received an endowment of 50 ECUs. In this task, subjects had to choose between accepting and rejecting a fifty-fifty bet. If a subject accepted the bet, they gained 50 ECUs on top of their endowment with probability 0.5 and lose $y \in \{10, 15, 20, 25, 30, 35, 40, 45, 50, 55\}$ ECUs from their endowment (and from their show-up fee, if necessary) with probability

0.5. If the bet was rejected, subjects simply kept their endowment. The subjects had 3 minutes to make all 10 choices. If a subject did not make a decision for a bet when the timer expired, they automatically rejected that bet. One out of the 10 choices got drawn by the computer completely at random to be payoff relevant for this task. This task's earnings were then determined according to the subject's choice and the state of the world.

The generosity task was derived from Crumpler and Grossman (2008). All subjects received an endowment of 100 ECUs for this task. Subjects had the opportunity to donate (some of) that endowment to the children's hospice Sterntaler in Mannheim. The donations that this non-profit organization receives are used to support families of (terminally) sick children and to organize activities for these children and families. The children's hospice was chosen as it is a local and relatively small scale uncontroversial charity. This means that small donations make a bigger relative impact on the charity's budget. Moreover, as the majority of subjects were in their early twenties, it is unlikely they would directly benefit from making a donation to this charity. To incentivize donations, the donations made by subjects were doubled. The donation was only made if the computer selected the donation task to be paid out to the subject. In total, 172 Euros were donated to the hospice.

The honesty task was derived from Fischbacher and Foellmi-Heusi (2013). Subjects had a die on their desk throughout the experiment. The task's instructions stated that they had to roll the die once and self-report the outcome. The subjects could do so without supervision. If a subject reported that they rolled six eyes, they received nothing for this task. Otherwise, their earnings for this task were 30 ECUs per self-reported eye rolled.

3.3.2 Part B: Productivity part

Part B only consisted of one task where subjects had to count the number of zeros in tables with 70 cells with binary values for 14 minutes (Abeler et al., 2011). The task is considered to be fair in terms of abilities across gender and subjects' strengths. This task was also chosen as subjects had to put in effort in order to generate money. If subjects submitted the correct number of zeros, they scored a point. The earnings for this task were 4 ECUs per point.

The number of zeros (i.e. the correct answer) was always between 15 and 55. The correct answer was drawn from a uniform distribution. This was done to prevent the existence of a most likely correct answer. If the value of every individual cell was independently drawn at random, then the binomial distribution would lead to 35 being four times as likely to be correct than under the uniform distribution. This could lead to some subjects always guessing 35 as answer instead of counting.

3.3.3 Part C: Social interactions part

While in Part A and B, decisions only had an influence on own payoffs and on the donation to the children's hospice. In this part, subjects influenced each others earnings. This part includes a competition game (Niederle and Versterlund, 2007) with a two-minute slider task (Gill and Prowse, 2012), a trust game (Berg et al., 1995), an ultimatum game (Güth et al., 1982) and a standard public goods game (Ledyard, 1995).

The competition game was similar to Niederle and Versterlund (2007), while the subjects did the slider task from Gill and Prowse (2012). The game was played in two rounds. In the first round, subjects had to compete against each other in random groups of four. Subjects had 2 minutes to place as many sliders to exactly the value 50, the middle value of each slider. For every slider on value 50, they scored 1 point. After the two minutes, subjects got feedback. If the subject scored better than their three competitors, their earnings were 8 ECUs per point. Otherwise, the earnings in the first round were 0 ECUs. In case of a tie at first place, the winner was selected at random.

For the second round, subjects got reallocated to a new group of four at random. Each subject had to make a decision whether to compete under a competition scheme (as in round 1) or in a piece rate payments scheme, where subjects got 2 ECUs per point regardless of the other players' scores. In the competition scheme, the subject had to compete against the scores of the three competitors from the first round. This prevented subjects from competing against a self selected sample of competitors. The final earnings for this task were the earnings from one of the two rounds drawn at random.

In the trust game (Berg et al., 1995), the session was split up into random pairs, where one person got role A (person A) and the other person got role B (person B) at random. Both persons received 100 ECUs as initial endowment for this task. Person A had the opportunity to send (some of) their endowment to person B. This amount was then tripled. Consequently, person B had the opportunity to send money back to person A. If person A decided to send x ECUs to person B and person B decided to send y ECUs to person A, then the earnings for this task were $100 - x + y$ ECUs for person A and $100 + 3x - y$ ECUs for person B.

In the ultimatum game (Güth et al., 1982), the session was again split up into new random pairs, where one person got role A (person A) and the other person got role B (person B) at random. Person A received 200 ECUs as initial endowment for this task. Person A had to make an offer between 0 and 200 ECUs to person B. Afterwards, person B had to either accept or reject person A's offer. If person A decided to offer x ECUs to person B and person B decided to accept the offer, then the earnings for this task were as $200 - x$ ECUs for person A and x ECUs for person B. If person B rejected the offer, both subjects earned 0 ECUs for this task.

The cooperation task was a public goods game in its most standard form (Ledyard, 1995). It was conducted in groups of four. It was an eight period game where every period subjects got an endowment of 10 ECUs on their personal account. Every subject could contribute an integer number of ECUs to the common account in every period. Everything a subject contributed to the common account got multiplied by 1.6 and got equally divided between the personal accounts of all four group members. At the end of every period, subjects received feedback on the aggregate contributions to the common account. The subject earnings for this task were the sum of all eight personal accounts at the end of a period.

3.3.4 Questionnaire

At the end of the experiment, subjects were asked about their age, gender, nationality and field of study. Moreover, the subjects were asked in a questionnaire on their well-being, tiredness, warmness, breathing and attitudes towards face masks on a five-point Likert-scale. The questions in the questionnaire were similar to Fikenzer et al. (2020) and the questions can be found in the Appendix. This questionnaire was done at the end of the session to prevent any experimental demand effect.

3.3.5 Experimental treatment

In the first seven sessions, between 23 September 2021 and 30 March 2022, 76 participants participated who all had to wear a medical face mask during the entire session. These sessions are considered to be the treatment sessions. There are many different type of masks one can wear and these masks can have different physiological effects (Fikenzer et al. (2020);Li et al. (2005)). During the first four sessions, both surgical and FFP2 masks were allowed. During the last three treatment sessions, only FFP2 masks were allowed. The participants were casually reminded if necessary that they had to keep wearing the mask throughout the experiment.

In the last seven sessions, between 5 April 2022 and 11 October 2022, 84 participants participated after the face masks mandate was abolished at the University of Mannheim. The participants were casually reminded if necessary that they could take off their mask once seated.

As the sample consists of mainly young healthy high educated individuals with an average age of 22, results are likely to be more conservative than with a representative sample as, for example, wearing a face mask can have a larger effect for people with respiratory difficulties (Kyung et al., 2020). Also, the treatment effects are slightly more conservative as a few people in the treatment sessions did not wear their mask properly all the time and a few participants in the no-mask sessions wore masks voluntarily during (parts of) the session. The treatment is the mask mandate of the state of Baden-Württemberg to which the vast majority of participants complied.

The experiment was executed in two different orders. In Order 1, subjects began with Part A (individual characteristics) and ended with Part C (social interactions). In Order 2, subjects began with Part C and ended with Part A. It was randomized whether a session had Order 1 or Order 2. This is to identify the time effect of wearing a mask. Part A and Part C started approximately 45 minutes from each other. Therefore, these two randomized orders can be used to see the effect of continuously wearing a mask for a relatively shorter or longer amount of time.

3.4 Results

To identify the effects of wearing a mask, we can simply compare the samples with a non-parametric test at first. Unless specified otherwise, the p-values from the Wilcoxon rank-sum test are used for statistical comparison of the samples. The results are shown in Table 3.1. There are 14 null hypotheses that are tested. Therefore, I use multiple hypothesis testing and also show the q-values to control the positive false discovery rate (pFDR) (Storey and Tibshirani, 2003). For p-values smaller than 0.05, Cohen's d is shown to measure the effect size.

It can be seen in Table 3.1 that wearing a mask does not affect individual characteristics or interactions in a negative way. There is a significant seven percentage point *increase* in ultimatum game offers when masks are worn. Also, there is no difference in productivity when doing the 14-minute counting task. However, there is a significant reduction in the productivity of a high intense two-minute slider task. In the first round, where everyone is competing against each other, this reduction is 9.1 percent. In the second

Table 3.1: Non-parametric treatment differences

			Treatment Differences		
	With Mask	Without Mask	Δ	q-value	Cohen's d
	(1)	(2)	(3)	(4)	(5)
Part A: Individual Characteristics					
Cognition score (share)	.763	.738	.025	.297	
	(.014)	(.013)	(.144)		
Risk aversion (switching point)	70.235	70.242	−.007	.869	
	(1.870)	(1.764)	(.981)		
Loss aversion (switching point)	27.683	30.711	−3.028**	.101	.277
	(1.267)	(1.180)	(.033)		
Donation (share)	.423	.414	.009	.639	
	(.042)	(.044)	(.515)		
Honesty (share)	.779	.814	−.035	.535	
	(.033)	(.030)	(.388)		
Part B: Productivity					
Counting Zeros (number of tables)	32.461	32.381	.080	.714	
	(.915)	(.873)	(.691)		
Part C: Interactions					
Slider: Scores round 1	21.079	22.988	−1.909***	.002	.501
	(.415)	(.435)	(<.001)		
Slider: Scores round 2	23.067	25.893	−2.826***	<.001	.663
	(.512)	(.443)	(<.001)		
Slider: Competition (share)	.434	.452	−.018	.781	
	(.057)	(.055)	(.819)		
Trust game: Sending (share)	.361	.381	−.020	.690	
	(.040)	(.043)	(.662)		
Trust game: Return (share of sending)	1.422	1.247	.175	.461	
	(.094)	(.098)	(.260)		
Ultimatum game: Offer (share)	.431	.361	.070**	.043	.649
	(.009)	(.014)	(.010)		
Ultimatum game: Acceptance (share)	.947	.881	.066	.467	
	(.026)	(.036)	(.301)		
Public good contribution (share)	.497	.403	.094	.458	
	(.057)	(.048)	(.255)		
Observations	76	84			

Notes: In column 1 and 2, the averages are shown of the groups with and without masks, respectively. The standard errors are between parentheses. In column 3, the treatment differences are shown. * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. p-values of the Wilcoxon rank sum test are displayed in brackets. In column 4, the q-values are shown to check for the positive false discovery rates. Column 5 shows the Cohen's d for the significant p-values to measure effect size.

round, the reduction is even larger (12.3 percent) which could be due to faster depletion when wearing a mask. The effect size is also relevant as

Cohen's d is .501 and .663 for the first and second round, respectively.

3.4.1 Productivity task with high intensity

As the only negative effect of wearing masks for real-effort tasks is on the number of correctly placed sliders, this section focuses on the slider task. The distribution of the slider scores can be seen in Figure 3.1. The score distribution with masks seems to be directly to the left of the distribution without masks.

The scores on session level are shown in Figure 3.2. It can be seen that the average session scores are lowest during sessions 5 to 7. This is when subjects had to wear FFP2 masks. During the first sessions, the average score is slightly higher, but still lower than the average scores of the control sessions.

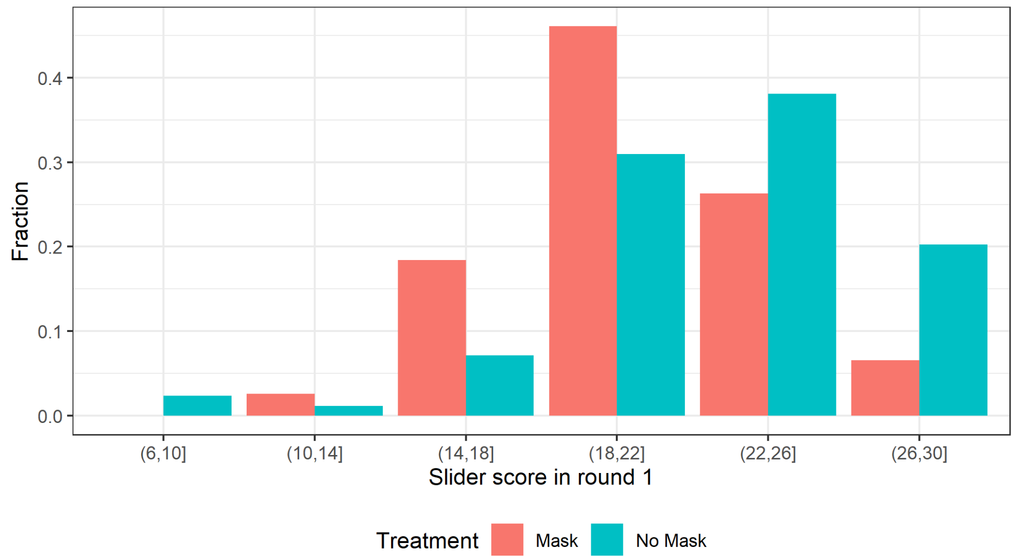
Figure 3.2 does not include session 4, as scores in that session are significantly lower than in other sessions. A similar Figure to Figure 3.2, but with session 4 included can be found in the Appendix. As the scores in this session are so much lower and this difference is not explicable, the fourth session is omitted from further analysis. One explanation is that there could have been a slower response from the server during that specific session. This could have increased the refresh time of the sliders and lowered the scores across-the-board.

As the scores in session 4 are a negative outlier and that session is a treatment session, the estimates of the remainder of the paper are more conservative than if the fourth session would be included. The samples are still significantly different across treatments when session 4 is omitted (p-values .033 and .016 for round 1 and 2, respectively).

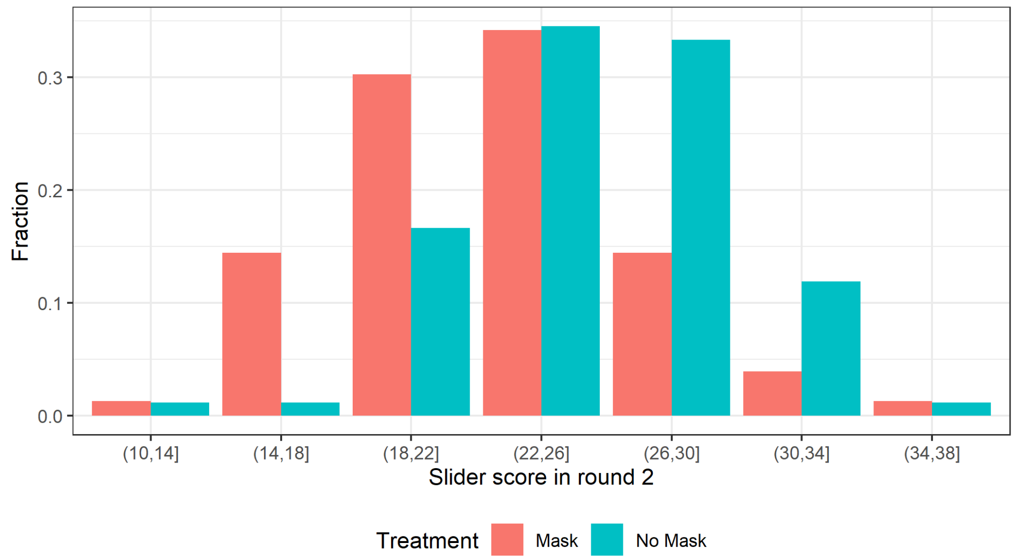
As the mask treatment followed the mask mandate of the state of Baden-Württemberg and was not imposed at random, it is important to compare the characteristics of the samples. The summary statistics are found in Table 3.2. It has to be noted that the sample is slightly different across treatment groups. The subjects in the no-mask treatment sessions were 10 months younger and statistically slightly more likely to be female and German. Therefore, it is important to control for these variables in linear regressions, to exclude the possibility that these differences (partially) drive the differences in the non-parametric sample comparisons in Table 3.1.

To include controls, we can look at the treatment effect in a standard multiple regression as in equation 3.1. In this equation, y is one of the dependent variables of interest, which are slider task scores in round 1 and round 2. M is the mask treatment which equals 1 if subject i had to wear

Figure 3.1: Distribution of slider scores by round and treatment



(a) Slider score distribution in round 1



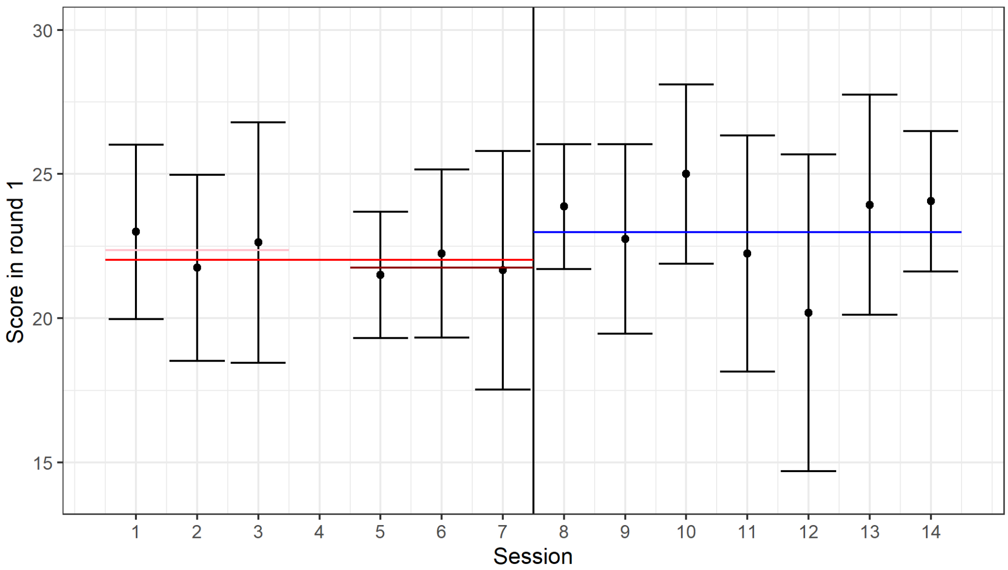
(b) Slider score distribution in round 2

a mask and 0 otherwise. X_i consists of individual controls as gender, age, nationality and field of study. Then, α_1 is the estimand that shows the effect of wearing a face mask on y_i.

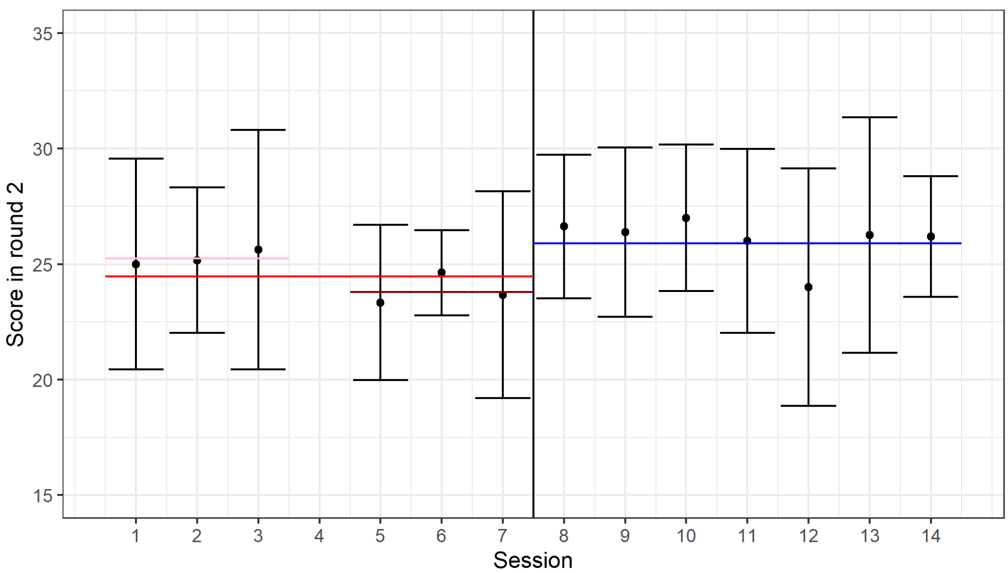
$$y_i = \alpha_0 + \alpha_1 * M_i + X_i * \gamma + \upsilon_i \tag{3.1}$$

The results of the regression can be seen in Table 3.3. It can be seen that controlling for individual characteristics amplifies the treatment effect

Figure 3.2: Averages of slider scores by session



(a) Slider score averages in round 1



(b) Slider score averages in round 2

Note: The y-axis in subfigures are the average number of correctly calibrated sliders. The blue lines show the averages of all individual scores without masks and the red lines show the averages of all individual scores with masks. The dark red lines are the averages across sessions with FFP2 masks and the lightred lines are the averages across all surgical/FFP2 mask sessions. The whiskers indicate the sessionwide standard deviations.

by approximately 30 percent. Wearing a mask lowers the score with 1.26 to 1.79 sliders, or approximately 5.6 to 7.1 percent. This is more evidence that the differences in output in the slider task are indeed caused by the masks and not by the sample differences. Beside the treatment effect, the

Table 3.2: Sample characteristics differences

	Summary statistics		
	With Mask	Without Mask	p-value
	(1)	(2)	(3)
Female	.461 (.502)	.595 (.494)	.090
Age	22.45 (2.62)	21.63 (3.18)	.015
German	.592 (.494)	.726 (.448)	.075
Economics	.263 (.443)	.273 (.448)	.882
Earnings	137.2 (48.2)	133.8 (57.3)	.458
Observations	76	84	

Notes: In column 1 and 2, the averages are shown of the groups with and without masks, respectively. The standard deviation is between parentheses. In column 3, the p-values of the Wilcoxon rank sum test are shown. effect size.

control variables show that women score higher than men, especially in the first round.

The question that remains is why the slider scores are affected by the treatment, unlike the output in the 14-minute counting task. It is likely that when a task lasts only two minutes, all subjects work under peak capacity which is likely to be reduced when wearing a mask similar to masks leadaing to a reduced maximal physical work capacity (Louhevaara, 1984). On the other hand, when subjects count zeros for 14 minutes, it is likely that subjects from both treatment groups did not perform under peak capacity and then a mask does not impose an effect on productivity at all.

Two possible channels that could enhance the treatment effect are the effects of wearing the mask on competition and depletion. First, it is shown that the lack of competition does not diminish the treatment effect. Afterwards, it is shown that the effect of wearing a mask is stronger when the slider task is in the second half of the experiment. This means that wearing a mask continuously for a longer period of time enhances the treatment effect.

3.4.2 Effect of competition

We have to consider the possibility that the treatment effect comes from the stress of the competition, which did not apply to the counting task which had a piece rate payment scheme. If competition was the driver of the difference, the treatment effect should disappear when looking at the scores of subjects with the piece rate payment scheme. However, the slider

Table 3.3: Treatment effect on the slider task scores

	Slider task scores			
	Score round 1		Score round 2	
	(1)	(2)	(3)	(4)
Mask Treatment	−.955	−1.258**	−1.426**	−1.786***
	(.600)	(.585)	(.657)	(.641)
Age		.159		.090
		(.105)		(.119)
German		−.347		−.093
		(.628)		(.696)
Female		2.762***		1.680***
		(.604)		(.621)
Economics		−.585		.023
		(.678)		(.738)
Session size		−.164*		−.173*
		(.087)		(.098)
Constant	22.988***	20.465***	25.893***	25.278***
	(.433)	(2.635)	(.441)	(3.039)
Observations	144	144	144	144
R^2	.016	.170	.031	.092

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The regressions in the even columns include control variables on the individual level.

scores were also negatively affected by the mask treatment when subjects were not facing competition. The productivity-loss is still significant when exclusively looking at scores under the piece rate payment scheme (p-value = .006). The distribution of scores under the piece rate payment scheme can be found in the Appendix.

Of course, the sample receiving a piece rate payment per slider is self selected. However, there is no difference in likelihood of entering competition between treatments. The competitive payment scheme itself does not significantly enhance the effect of wearing a face mask in round 2 as can be seen in Table 3.4. The subjects who chose the piece rate payments scheme (56 percent of all subjects) were also significantly affected by wearing a face mask and the effect size is statistically similar to the effect on the full sample.

Table 3.4: Slider task score by payment scheme

	Score round 2		
	Piece rate	Competitive	All schemes
	(1)	(2)	(3)
Mask Treatment	-1.869^{**}	-2.133^{**}	-1.522^{**}
	(.802)	(.910)	(.775)
Competitive scheme			2.899^{***}
			(.848)
Mask x Competitive			$-.410$
			(1.256)
Constant	25.693^{***}	26.901^{***}	24.819^{***}
	(3.943)	(4.706)	(2.846)
Observations	80	64	144
Control variables	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust.

3.4.3 Effect of timing

A potential channel that lowers scores in treatment sessions is tiredness. This could also explain that the treatment effect is larger in the second round than in the first round. We can use the randomized order to see if subjects scored worse if the slider task was in the second half of the experiment due to wearing the masks for a longer period of time. It is interesting to compare the subjects who had to start with part C to those who started with part A. The people who started with part A, had to do the slider task in the second half of the experiment, which was 45 minutes later than those who started with part C.

We can exploit this randomized variation to estimate whether the effect of wearing a face mask is consistent over time or not. To get the time effect easily, a difference-in-differences (DID) analysis is used. The DID equation 3.2 allows to identify the differences over time. L equals 1 if the task is done by subject i later and 0 if the task is done earlier. The β_3 estimator will show the difference between wearing a mask or not while assuming similar time trend.

$$y_i = \beta_0 + \beta_1 * M_i + \beta_2 * L_i + \beta_3 * M_i * L_i + X_i * \delta + \epsilon_i \tag{3.2}$$

The timing does matter as can be seen in Table 3.5. Having worn a mask longer does significantly affect the treatment effect in round 1 compared to

Table 3.5: Treatment effect on the slider task pver time

	Slider task scores			
	(1)	(2)	(3)	(4)
Panel A: Score round 1				
Mask Treatment	-.955	-1.258^{**}	-2.259^{***}	-2.536^{***}
	(.600)	(.585)	(.636)	(.645)
Later			2.701^{***}	3.069^{***}
			(.875)	(.798)
Mask x Later			-3.253^{**}	-2.596^{**}
			(1.304)	(1.112)
Panel B: Score round 2				
Mask	-1.426^{**}	-1.786^{***}	-2.189^{***}	-2.434^{***}
	(.657)	(.641)	(.775)	(.761)
Later			1.222	1.386
			(.905)	(.863)
Mask x Later			-2.120	-1.548
			(1.410)	(1.371)
Observations	144	144	144	144
Control variables	NO	YES	NO	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The regressions in the even columns include control variables on the individual level.

only wearing it for a few minutes. Without a mask, it is actually beneficial to the score if the task comes later in the experiment. This might be because the slider task is the first task in the first half of the experiment when subjects were possibly not yet fully accustomed to the experiment. This is plausible as most of this benefit disappeared in second round.

However, the benefit from being used to the experiment fully disappeared when wearing a mask where scores in the first round are approximately three points lower than when subjects without masks did the task in the second half of the experiment. In round 2, the interaction effect is weaker and statistically no longer significant across all payment-schemes, even though the signs are still the same.

3.4.4 Other tasks

The regression estimates of the treatment effect (over time) are shown in Table 3.6. This Table shows results from regressions both with and without controlling for the individual subject characteristics. The first two columns look at α_1 from equation 3.1 and the last two columns look at estimations

Table 3.6: Treatment effect on the other tasks: Regression estimates

	Estimate			
	Mask treatment		Mask x Later	
	(1)	(2)	(3)	(4)
Part A: Individual Characteristics				
Cognition	.025	.016	-.009	-.008
	(.019)	(.019)	(.039)	(.038)
Risk aversion (switching point)	-.006	-1.825	5.392	6.870
	(2.554)	(2.376)	(5.065)	(4.764)
Loss aversion (switching point)	-3.028*	-1.078	-2.155	-3.223
	(1.731)	(1.561)	(3.515)	(3.153)
Donation	.009	.017	.153	.161
	(.061)	(.061)	(.123)	(.121)
Honesty	-.035	-.022	-.055	-.064
	(.044)	(.043)	(.085)	(.082)
Part B: Productivity				
Counting Zeros	.080	-.008		
	(1.256)	(1.341)		
Part C: Interactions				
Slider: Competition	-.018	.006	-.072	-.052
	(.079)	(.076)	(.159)	(.143)
Trust game: Sending	-.020	.009	-.118	-.082
	(.082)	(.082)	(.159)	(.153)
Trust game: Return	-.004	.082	.138	.149
	(.167)	(.155)	(.339)	(.343)
Ultimatum game: Offer	.070***	.099***	.004	.004
	(.024)	(.024)	(.045)	(.043)
Ultimatum game: Acceptance	.066	.095	.128	.140
	(.062)	(.073)	(.120)	(.113)
Public good: Contributions	.094**	.118	-.054	-.054
	(.044)	(.073)	(.086)	(.083)
Observations	144	144	144	144
Control variables	NO	YES	NO	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The regressions in the even columns include control variables on the individual level.

for β_1 from equation 3.2.

It can be seen that the masks only have a significant positive effect on the offer in the ultimatum game. This effect is similar to the difference in the non-parametric sample comparisons. This implies that in session where masks were worn, there were higher offers in the ultimatum game. This

Figure 3.3: Distribution of ultimatum game offers

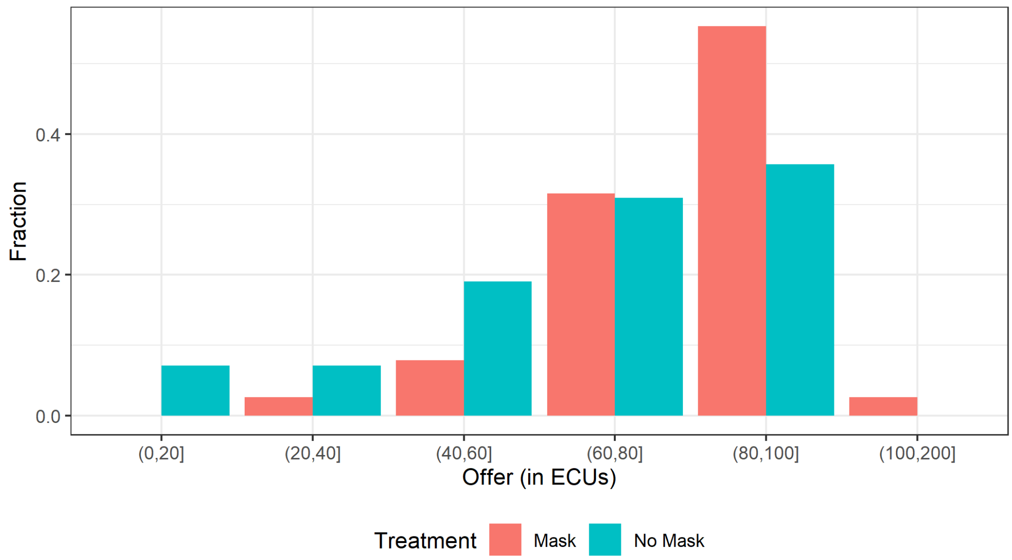
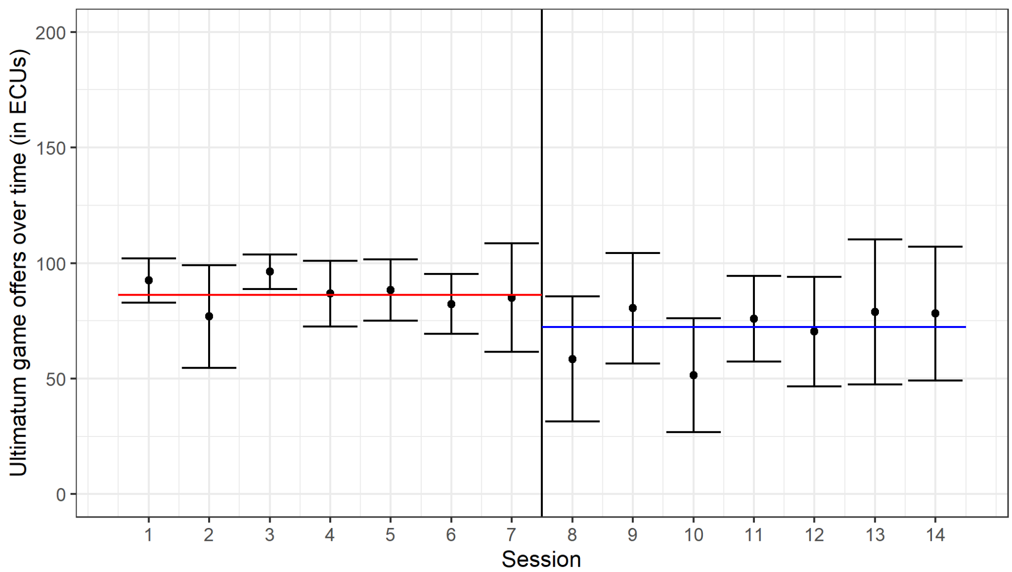


Figure 3.4: Average ultimatum game offers over time



Note: The y-axis shows the average offer on session level. The blue line shows the averages of all individual offers without masks and the red line shows the averages of all individual scores with masks. The whiskers indicate the sessionwide standard deviations.

result is robust when controlling for individual characteristics and does not depend on the duration of wearing the mask. The distribution of the offers is shown in Figure 3.3 and the average offers over sessions are displayed in Figure 3.4. It can be seen that the offers in the ultimatum game dropped immediately after the mask mandate was abolished. Sessions 7 and 8 were only six days apart.

The treatment effect is sizable as the average offer in the mask-less ses-

Table 3.7: Ultimatum game offers

	Dependent variable: Ultimatum game offer			
	(1)	(2)	(3)	(4)
Mask Treatment	.070***	.079***	.072**	.075**
	(.024)	(.023)	(.035)	(.031)
Later			-.019	-.015
			(.038)	(.032)
Mask x Later			.004	-.007
			(.045)	(.038)
Public good contribution		.115***		.118***
		(.036)		(.037)
Donation		.025		.025
		(.024)		(.024)
Dice outcome		.016		.012
		(.052)		(.051)
Trust game sending		.031**		.032**
		(.014)		(.015)
Constant	.361***	.454***	.353***	.446***
	(.020)	(.089)	(.029)	(.086)
Observations	80	80	80	80
Control variables	NO	YES	NO	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The regressions in the even columns include control variables on the individual level.

sions is 0.65 standard deviations lower as seen in Table 3.1. It is unlikely that offers in the treatment sessions are higher due to generosity, as there are no other outcomes from other tasks that are significantly affected by wearing a mask, including the donation task, trust game and public good game. Hence, even when controlling for the choices in these tasks in Table 3.7, the effect of wearing a mask remains almost identical. There is positive correlation between the offers in the ultimatum game and cooperation in the trust and public good games, but the higher offers are not due to more generous or cooperative proposers in the mask treatment.

It could be that the offers are higher due to the expectation that a lower offer would be rejected by the recipient. Offers in the ultimatum game are more likely to be accepted if the proposer has a positive facial expression (Mussel et al., 2013). As there are no facial expressions when people wear a mask in the lab, proposers might have believed that they had to increase

Table 3.8: Conditional acceptance of ultimatum game offers

	Dependent variable:			
	Ultimatum game acceptance			
	(1)	(2)	(3)	(4)
Mask Treatment	.066	.064	-.223	-.228
	(.062)	(.079)	(.360)	(.356)
Offer		.361	.227	.196
		(.341)	(.423)	(.404)
Mask x Offer			.635	.703
			(.792)	(.792)
Constant	.881***	1.087***	.799***	1.069***
	(.050)	(.367)	(.169)	(.373)
Observations	80	80	80	80
Control variables	NO	YES	NO	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. The regressions in the even columns include control variables on the individual level.

their offer to get the recipient to accept it. This would also explain why there is just an effect from wearing a mask and not from the duration of wearing a mask or from the type of mask worn. This would be a very subconscious effect, as it is unlikely that proposers were aware of the finding from Mussel et al. (2013). Moreover, both the higher offer and the lower expected acceptance rate are not in line with rational choice.

However, does wearing a mask actually matter for the acceptance of the offer? According to the sample comparisons, receivers are not less likely to accept the offer. As the offers are higher, the question is whether the conditional acceptance rate is lower when wearing a mask. In Table 3.8, there is no evidence that wearing a mask leads to lower acceptance rates from offer receivers. This means that if there are different beliefs about acceptance, that these beliefs are unjustified.

3.5 Concluding remarks

This paper examines the effect of face masks on economic behavior and productivity. Wearing a mask does not lead to differences in cognition, risk aversion, honesty or generosity. Also wearing a mask does not lead to more antisocial behavior. Offers in an ultimatum game are around 20 percent higher. This is not due to mask wearers being more generous or having different social preferences as there are no effect of masks on other social interaction tasks. Most likely, offers are higher due to incorrect different beliefs in acceptance thresholds of receivers. However, further research would be required to confirm this.

There is a significant difference in productivity under a high intensity task. When subjects had to put in maximal effort for two minutes, masks lead to 5.6 to 7.1 percent lower productivity. This effect is amplified if a mask is worn for 45 minutes longer. The effect is independent of the payment scheme. When people can pace their efforts during a 14-minute task with a piece rate payment scheme, the mask has no effect on performance. This implies that there is a difference in the effect of wearing a mask during productivity tasks with lower and higher intensity.

Therefore, when deciding on wearing a mask or on (re-)introducing a mask mandate, it has to be taken into account that there are negative side effects of wearing a mask in specific cases where one has to perform under peak capacity. However, there are otherwise no negative effects of wearing a mask and masks do not appear to affect people in cognition, economic preferences or social interactions.

References

Hoda M. Abd-Elfattah, Faten H. Abdelazeim, and Shorouk Elshennawy. Physical and cognitive consequences of fatigue: A review. *Journal of Advanced Research 6 (3)*, pages 351–358, 2015.

Johannes Abeler, Armin Falk, Lorenz Goette, and David Huffman. Reference points and effort provision. *American Economic Review 101 (2)*, pages 470–492, 2011.

Ran Abramitzky, Victor Lavy, and Maayan Segev. The effect of changes in the skill premium on college degree attainment and the choice of major. *NBER Working Papers 26420*, 2019.

Tweede Fase Adviespunt. Zakboek tweede fase (in dutch). Technical report, Tweede Fase Adviespunt, 2007.

Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R Thomas. Equilibrium grade inflation with implications for female interest in stem majors. Working Paper 26556, National Bureau of Economic Research, December 2019.

Steffen Altmann, Armin Falk, Paul Heidhues, Rajshri Jayaraman, and Marrit Teirlinck. Defaults and donations: Evidence from a field experiment. *Review of Economics and Statistics 101 (5)*, pages 808–826, 2019.

Joseph Altonji, Peter Ardiciacono, and Arnaud Maurel. The analysis of field choice in college and graduate school: Determinants and wage effects. *The Handbook of the Economics of Education volume 5*, pages 305–396, 2016.

James Archsmith, Anthony Heyes, and Soodeh Saberian. Air quality and error quantity: Pollution and performance in a high-skilled, quality-focused occupation. *Journal of the Association of Environmental and Resource Economists, 5 (4)*, pages 827–863, 2018.

Peter Ardiciacono. Ability sorting and the returns to college major. *Journal of Econometrics 121*, pages 343–375, 2004.

Esteban Aucejo and Jonathan James. The path to college education: The role of math and verbal skills. *Journal of Political Economy 129*, pages 2905–2946, 2021.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity and social history. *Games and Economic Behavior 10*, pages 122–142, 1995.

Joseph Bertrand. *Review of Theorie mathematique de la richesse sociale and of Recherches sur les principles mathematiques de la theorie des richesses.* Journal des Savants 67, 1883.

Eric Bettinger and Bridget Long. Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review 95*, pages 152–157, 2005.

Martin Biewen and Jakob Schwerter. Does more math in high school increase the share of female stem workers? evidence from a curriculum reform. IZA Discussion Paper No12236, March 2019.

Paul Bingley, Eskil Heinesen, Karl Fritjof Krassel, and Nicolai Kristensen. The timing of instruction time: Accumulated hours, timing and pupil achievement. IZA Discussion Paper No11807, September 2018.

BLS. Employment projections program. Technical report, Bureau of Labor Statistics, 2021.

Thomas Breda, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre. Do female role models reduce the gender gap in science? evidence from french high schools. IZA Discussion Paper 13163, April 2020.

Scott Carrell, Marianne Page, and James West. Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*, pages 1101–1144, 2010.

CDC. Recommendation regarding the use of cloth face coverings, especially in areas of significant community-based transmission, 2020.

Cedefop. Cedefop skills forecast. Technical report, European Commission. Bruxelles., 2020.

Gary Charness, Uri Gneezy, and Brianna Halladay. Experimental methods: Pay one or pay all. *Journal of Economic Behavior and Organization 131*, pages 141–150, 2016.

Daniel Chen, Martin Schonger, and Chris Wickens. otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, pages 88–97, 2016.

Jiu-Chiuan Chen and Joel Schwartz. Neurobehavioral effects of ambient air pollution on cognitive performance in us adults. *NeuroToxicology 30 (2)*, pages 231–239, 2009.

Xianglei Chen and Matthew Soldner. Stem attrition: College students' paths into and out of stem fields. *National Center for Education Statistics (1)*, 2014.

Pierre-André Chiappori and Monica Paiella. Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association 9 (6)*, pages 1021–1052, 2011.

Angela Cools, Raquel Fernandez, and Eleonora Patacchini. Girls, boys and high achievers. 2019. IZA Discussion Papers 12314, Institute of Labor Economics (IZA).

Kalena Cortes, Joshua Goodman, and Takako Nomi. Intensive math instruction and educational attainment: Long- run impacts of double- dose algebra. *Journal of Human Resources 50*, pages 108–158, 2015.

Heidi Crumpler and Philip J. Grossman. An experimental test of warm glow giving. *Journal of Public Economics 92*, pages 1011–1021, 2008.

Michael R. Cunningham. Weather, mood, and helping behavior: Quasi experiments with the sunshine samaritan. *Journal of Personality and Social Psychology 37 (11)*, pages 1947–1956, 1979.

Algemeen Dagblad. Minder mensen stapten over naar andere zorgverzekeraar (in dutch), 2020. https://www.ad.nl/binnenland/minder-mensen-stapten-over-naar-andere-zorgverzekeraar ae858d83/.

Marta De Philippis. Stem graduates and secondary school curriculum: Does early exposure to science matter? *Journal of Human Resources*, 2021.

Judith M. Delaney and Paul J. Devereux. Understanding gender differences in stem: Evidence from college applications. *Economics of Education Review 72*, pages 219–238, 2019.

Ap Dijksterhuis, Ad van Knippenberg, Arie W. Kruglanski, and Carel Schaper. Motivated social cognition: Need for closure effects on memory and judgment. *Journal of Experimental Social Psychology 32 (3)*, pages 254–270, 1996.

Raoul du Pre. Meteen na de invoering ging het al mis (article in dutch). *De Volkskrant, October 20*, 2005.

M.L. Duch, M.R.P. Grossmann, and T. Lauer. z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance, Volume 28*, page Article 100400, 2020.

Martin Dufwenberg and Uri Gneezy. Price competition and market concentration: an experimental study. *International Journal of Industrial Organization vol 18(1)*, pages 7–22, 2000.

DUO. Jaarverslag numerus fixi. studiejaar 2010-2011 (in dutch). Technical report, Dienst Uitvoering Onderwijs, 2011.

ECDC. Using face masks in the community-reducing covid-19 transmission from potentially asymptomatic or presymptomatic people through the use of face masks, 2020.

Christoph Engel and Oliver Kirchkamp. How to deal with inconsistent choices on multiple price lists. *Journal of Economic Behavior and Organization 160*, pages 138–157, 2019.

Birte Englich and Kirsten Soder. Moody experts - how mood and expertise influence judgmental anchoring. *Judgment and Decision Making 4 (1)*, pages 41–50, 2009.

Helga Fehr-Duda and Thomas Epper. Probability and risk: Foundations and economic implications of probability-dependent risk preferences. *Annual Review of Economics 2012 (4)*, pages 567–593, 2012.

Leslie K. Ferrarelli. Hunger signals suppress risk perception. *Science Signalling 9 (458)*, pages 292–293, 2016.

Sven Fikenzer, T. Uhe, D. Lavall, U. Rudolph, R. Falz, M. Busse, P. Hepp, and U. Laufs. Effects of surgical and ffp2/n95 face masks on cardiopulmonary exercise capacity. *Clinical Research in Cardiology (2020)*, pages 1–9, 2020.

Urs Fischbacher. z-tree: Zurich toolbox for ready–made economic experiments. *Experimental Economics 10*, pages 171–178, 2007.

Urs Fischbacher and Franziska Foellmi-Heusi. Lies in disguise: an experimental study on cheating. *Journal of the European Economic Association, 11(3)*, pages 525–547, 2013.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, pages 25–42, 2005.

Simon Gaechter, Eric J. Johnson, and Andreas Herrmann. Individual-level loss aversion in riskless and risky choices. Discussion Papers 2010-20, The Centre for Decision Research and Experimental Economics, School of Economics, University of Nottingham., 2010.

Michael Gerfin and Martin Schellhorn. Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits - swiss evidence. *Health Economics 15*, pages 1011–1020, 2006.

David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review, 102*, pages 469–503, 2012.

Joshua Goodman. The labor of division: Returns to compulsory high school math coursework. *Journal of Labor Economics 37 (4)*, pages 1141–1182, 2019.

Ben Greiner. Subject pool recruitment procedures: Organizing experiments with orsee. *Journal of the Economic Science Association, 1(1)*, pages 114–125, 2015.

Amanda L. Griffith. Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review 29*, page 911–922, 2010.

Werner Güth, Rolf Schmittberger, and Bernd Schwarze. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization 3*, pages 367–388, 1982.

Nina Guyon and Elise Huillery. Biased aspirations and social inequality at school: Evidence from french teenagers. *Economic Journal, Royal Economic Society, vol. 131 (634)*, pages 745–796, 2021.

Glenn W. Harrison and E. Elisabet Rutström. Risk aversion in the laboratory. *Research in Experimental Economics (12)*, page 41–196, 2008.

Anthony Heyes and Soodeh Saberian. Temperature and decisions: Evidence from 207,000 court cases. *American Economic Journal: Applied Economics 2019, 11(2)*, pages 238–265, 2019.

Charles Holt and Susan Laury. Risk aversion and incentive effects. *American Economic Review, 92 (5)*, pages 1644–1655, 2002.

Harold Hotelling. Stability in competition. *The Economic Journal 39*, pages 41–57, 1929.

ISCED. International standard classification of education. Technical report, UNESCO Institute for Statistics, 2011.

Estelle James, Nabeel Alsalam, Joseph Conaty, and Duc-Le To. College quality and future earnings: where should you send your child to college? *American Economic Review 79*, pages 247–252, 1989.

Juanna Joensen and Helena Nielsen. Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources 44 (1)*, pages 171–198, 2009.

Juanna Joensen and Helena Nielsen. Mathematics and gender: Heterogeneity in causes and consequences. *Economic Journal 126*, pages 1129–1163, 2016.

Arthur T. Johnson. Respirator masks protect health but impact performance: a review. *Journal of Biological Engineering, 10*, pages 525–547, 2016.

Moshe Justman and Susan Méndez. Gendered choices of stem subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of Education Review 64*, pages 282–297, 2018.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica 47 (2)*, pages 263–291, 1979.

Kanav Kahol, Mario J. Leyba, Mary Deka, Vikram Deka, Stephanie Mayes, Marshall Smith, John J. Ferrara, and Sethuraman Panchanathan. Effect of fatigue on psychomotor and cognitive skills. *American Journal of Surgery 195 (2)*, pages 195–204, 2008.

Paul Klemperer. Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Review of Economic Studies 62 (4)*, pages 515–539, 1995.

Edward Kokkelenberg and Esha Sinha. Who succeeds in stem studies? an analysis of binghamton university undergraduate students. *Economics of Education Review 29*, pages 935–946, 2010.

Miriam Krieger and Stefan Felder. Can decision biases improve insurance outcomes? an experiment on status quo bias in health insurance choice. *International Journal of Environmental Research and Public Health 10*, pages 2560–2577, 2013.

S.Y. Kyung, Y. Kim, H. Hwang, J.W. Park, and S.H. Jeong. Risks of n95 face mask use in subjects with copd. *Respiratory Care 65(5)*, pages 658–664, 2020.

Victor Lavy. Do differences in schools' instruction time explain international achievement gaps? evidence from developed and developing countries. *The Economic Journal, 125 (November)*, pages F397–F424, 2015.

Victor Lavy. Expanding school resources and increasing time on task: Effects of a policy experiment in israel on student academic achievement and behavior. *Journal of the European Economic Association*, February 2020.

John Ledyard. Public goods: A survey of experimental research. Technical report, University Library of Munich, Germany, 1995.

Nancy H. L. Leung, Daniel K. W. Chu, and Eunice Y. C. Shiu et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature Medicine volume 26*, pages 676–680, 2020.

Haim Levy. Absolute and relative risk aversion: An experimental study. *Journal of Risk and Uncertainty 8*, pages 289–307, 1994.

Y. Li, H. Tokura, Y.P. Guo amd A.S.W. Wong, T. Wong, J. Chung, and E. Newton. Effects of wearing n95 and surgical facemasks on heart rate, thermal stress and subjective sensations. *International Archives of Occupational and Environmental Health 78*, pages 501–509, 2005.

Veikko A Louhevaara. Physiological effects associated with the use of respiratory protective devices. *Scandinavian Journal of Work Environment and Health 10 (5)*, pages 658–664, 1984.

Wei Lyu and George L. Wehby. Community use of face masks and covid-19: Evidence from a natural experiment of state mandates in the us. *Health Affairs 38:3*, pages 1419–1425, 2020.

Anandi Mani, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. Poverty impedes cognitive function. *Science 341 (6149)*, pages 976–980, 2013.

Yusufcan Masatlioglu and Efe Ok. Rational choice with status quo bias. *Journal of Economic Theory*, pages 1–29, 2005.

Timo Mitze, Reinhold Kosfeld, Johannes Rode, and Klaus Wälde. Face masks considerably reduce covid-19 cases in germany. *Proceedings of the National Academy of Sciences 117 (51)*, pages 32293–32301, 2020.

Louis-Philippe Morin. Do men and women respond differently to competition? evidence from a major education reform. *Journal of Labor Economics 33 (2)*, pages 443–491, 2015.

Pierre Mouganie and Yaojing Wang. High-performing peers and female stem choices in school. IZA Discussion Paper No12455, June 2019.

Patrick Mussel, Anja S. Göritz, and Johannes Hewig. The value of a smile: Facial expression affects ultimatum-game responses. *Judgment and Decision Making, Vol. 8, No. 3*, pages 381–385, 2013.

Muriel Niederle and Lise Versterlund. Do women shy away from competition? do men compete too much? *Quarterly Journal of Economics (3)*, pages 1067–1101, 2007.

OECD. Oecd information technology outlook, 2010.

Mark V. Pauly. Dumb decisions or as smart as the average politician? : economic and behavioral explanations for insurance demand. In Erwann Michel-Kerjan and Paul Slovic, editors, *The Irrational Economist*, chapter 16, pages 200–215. Public Affairs, New York, 2010.

Rob Pijpers, Raoul Oudejans, and Frank Bakkerl. Changes in the perception of action possibilities while climbing to fatigue on a climbing wall. *Journal of Sports Sciences 25*, pages 97–110, 2007.

Franzis Preckel, Thomas Goetz, Reinhard Pekrun, and Michael Kleine. Gender differences in gifted and average-ability students: Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted Child Quarterly 52 (2)*, pages 146–159, 2008.

Caterina Primi, Kinga Morsanyi, Francesca Chiesi, Maria Anna Donati, and Jayne Hamilton. The development and testing of a new version of the cognitive reflection test applying item response theory (irt). *Journal of Behavioral Decision Making*, pages 1–17, 2015.

Holger Rau. The disposition effect in team investment decisions: Experimental evidence. *Journal of Banking and Finance 61*, pages 272–282, 2015.

John Raven. Raven progressive matrices. In McCallum R.S., editor, *Handbook of Nonverbal Assessment*, chapter 11, pages 223–237. Springer, Boston, MA, 2003.

Ilana Ritov and Jonathan Baron. Status-quo and omission biases. *Journal of Risk and Uncertainty 5*, pages 49–61, 1992.

RIVM. Inhoudelijke onderbouwing met betrekking tot adviezen over toepassing van niet-medische mondneusmaskers in openbare ruimten (in dutch), 2020.

RKI. Mund-nasen-bedeckung im öffentlichen raum als weitere komponente zur reduktion der Übertragungen von covid-19. strategie-ergänzung zu empfohlenen infektionsschutzmaßnahmen und zielen (3. update). (in german), 2020.

Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The Quarterly Journal of Economics 90 (4)*, pages 629–649, 1976.

William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty 1*, pages 7–59, 1988.

Anne Schlegtendal, Lynn Eitner, Michael Falkenstein, Anna Hoffmann, Thomas Lücke, Kathrin Sinningen, and Folke Brinkmann 1. To mask or not to mask—evaluation of cognitive performance in children - wearing face masks during school lessons (maskids). *Children 2022, 9, 95*, pages 1–11, 2021.

Dan Simon. *In Doubt: The Psychology of the Criminal Justice Process.* Harvard University Press, Massachusetts, 2012.

Edward Hugh Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Ser. B, 13,*, pages 238–241, 1951.

Mark Spils. Effect of quality of mathematics education on college major choices. a fuzzy difference-in-discontinuities design. Master's thesis, University of Mannheim, June 2018.

Education Statistics. Digest of education statistics. 53rd edition. Technical report, U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, 2017.

John D. Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences 100 (16)*, pages 9440–9445, 2003.

Bruce A. Strombom, Thomas C. Buchmueller, and Paul J. Feldstein. Switching costs, price sensitivity and health plan choice. *Journal of Health Economics 21*, pages 89–116, 2002.

Nadine Tchen, Helen G. Juffs, Fiona P. Downie, Qi-Long Yi, Hanxian Hu, Irene Chemerynsky, Mark Clemons, and et al. Cognitive function, fatigue, and menopausal symptoms in women receiving adjuvant chemotherapy for breast cancer. *Journal of Clinical Oncology 21 (22)*, pages 4175–4183, 2003.

Jean Tirole. *Theory of Industrial Organization*. The MIT Press, 1988.

Jeen van Hijum. Nederlandse zorgautoriteit: Monitor zorgverzekeringen 2021, 2021. https://magazines.nza.nl/nza-magazines/2021/01/monitor-zorgverzekeringen-2021.

K.P.M van Winssen, van Kleef R.C., and W.P.M.M. van de Ven. How profitable is a voluntary deductible in health insurance for the consumer? *Health Policy 119*, pages 688–695, 2015.

K.P.M van Winssen, van Kleef R.C., and W.P.M.M. van de Ven. Potential determinants of deductible uptake in health insurance: How to increase uptake in the netherlands? *European Journal Health Economics 17*, pages 1059–1072, 2016.

Russell M. Viner, Charlotte Clark, Stephanie J.C. Taylor, Kam Bhui, Emily Klineberg, Jenny Head, Robert Booy, and Stephen A. Stansfeld. Longitudinal risk factors for persistent fatigue in adolescents. *Archives of Pediatrics and Adolescent Medicine 162 (5)*, pages 469–475, 2008.

Leslie Jo Weaver and Craig Hadley. Moving beyond hunger and nutrition: A systematic review of the evidence linking food insecurity and mental health in developing countries. *Ecology of Food and Nutrition 48 (4)*, pages 263–284, 2009.

Linda Weinreb, Cheryl Wehler, Jennifer Perloff, Richard Scott, David Hosmer, Linda Sagor, and Craig Gundersen. Hunger: Its impact on children's health and mental health. *Pediatrics 110 (4)*, pages 41–50, 2002.

Wet-WVO. Wet wijzigingswet op het voortgezet onderwijs (aanpassing profielen tweede fase vwo en havo), 2006. https://wetten.overheid.nl/BWBR0019799/2007-08-01.

Mary Kay White, Thomas K. Hodous, and Max Vercruyssen. Effects of thermal environment and chemical protective clothing on work tolerance, physiological responses, and subjective ratings. *Ergonomics 34 (4)*, pages 445–457, 1991.

WHO. Advice on the use of masks in the context of covid-19, 2020.

Robert S. Wyer and Donal E. Carlston. *Social Cognition, Inference, and Attribution*. Lawrence Erlbaum Associates, New Jersey, 1979.

Jingyi Xiao, Eunice Y. C. Shiu, Huizhi Gao, Jessica Y. Wong, Min W. Fong, Sukhyun Ryu, and Benjamin J. Cowling. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings - personal protective and environmental measures. *Emerging Infectious Diseases 26 (5)*, pages 967–975, 2020.

Basit Zafar. College major choice and the gender gap. *The Journal of Human Resources Vol. 48 (3)*, pages 545–595, 2013.

Zorgverzekeringswet. Zorgverzekeringswet, 2005. https://wetten.overheid.nl/BWBR0018450/2022-07-01.

Appendices

A Appendix to Chapter 1

A.1 Additional Tables

Table A.1: STEM fields according to ISCED (2011)

STEM Fields	Non-STEM Fields
Sciences (life, physical, and computing) Technology (manufacturing, and processing) Engineering (including engineering trades and civil engineering) Mathematics (including operations research, numerical analysis, actuarial science, and statistics)	All other fields

Table A.2: Secondary school before and after the policy change

	HAVO			**VWO**		
	Old	New	Δ	Old	New	Δ
Compulsory courses						
Dutch	400	400	0	480	480	0
English	360	360	0	400	400	0
Third Language	160	0	-160	320	480	+160
Other (P.E. etc)	560	360	-200	760	560	-200
Nature/Tech						
Physics	440	400	-40	560	480	-80
Chemistry	280	320	+40	520	440	-80
Math B	440	360	-80	760	600	-160
Electives	560	1,000	+440	1,000	1,360	+360
Other fields						
Field courses	1,160	1,040	-120	1,840	1,440	-400
Electives	560	1,040	+480	1,000	1,440	+440
Total	3,200	3,200	0	4,800	4,800	0

Table A.3: Population long run outcomes

	VWO			HAVO		
	Younger Mean (SD)	Older Mean (SD)	Diff (p-value)	Younger Mean (SD)	Older Mean (SD)	Diff (p-value)
Nature/Tech	.320	.173	.147	.152	.111	.041
	(.467)	(.378)	(<.001)	(.359)	(.314)	(<.001)
Bachelor	.854	.853	.001	.653	.653	<.001
	(.353)	(.354)	(.543)	(.476)	(.476)	(.572)
STEM Bachelor	.157	.154	.003	.029	.029	<.001
	(.364)	(.361)	(.136)	(.168)	(.167)	(.671)
Master	.486	.477	.009	.101	.103	−.002
	(.5)	(.499)	(.038)	(.302)	(.303)	(.973)
STEM Master	.113	.111	.002	.021	.020	.001
	(.317)	(.314)	(.253)	(.142)	(.141)	(.394)
PhD	.066	.071	−.005	.002	.002	<.001
	(.248)	(.257)	(.007)	(.041)	(.041)	(.671)
Income	9.95	9.907	.043	10	9.976	.024
	(1.954)	(1.976)	(<.001)	(1.639)	(1.638)	(<.001)
Partner	.797	.794	.003	.75	.758	−.008
	(.402)	(.404)	(.564)	(.433)	(.423)	(.002)
Spouse	.131	.141	−.01	.188	.202	−.014
	(.337)	(.349)	(.052)	(.391)	(.402)	(.003)
Child(ren)	.085	.092	−.007	.168	.176	−.008
	(.279)	(.289)	(.178)	(.374)	(.381)	(.863)
Observations	38,191	37,625		46,264	45,701	

Table A.4: Heterogeneous effects by migration background

	Degree Completion					
	Nature/Tech		STEM Bachelor		STEM Master	
	Main	Placebo	Main	Placebo	Main	Placebo
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All students						
Treatment	.108***	−.001	.005	.0004	.002	−.003
	(.004)	(.003)	(.003)	(.003)	(.003)	(.003)
Observations	154,042	141,719	154,042	141,719	154,042	141,719
Panel D: By migration background						
Treatment **with migration background**	.113***	−.002	.001	−.004	.003	−.006
	(.009)	(.008)	(.008)	(.009)	(.007)	(.008)
Observations	24,396	22,459	24,396	22,459	24,396	22,459
Treatment **without migration background**	.107***	−.0004	.006*	.002	.002	−.002
	(.004)	(.004)	(.003)	(.004)	(.003)	(.003)
Observations	129,646	119,260	129,646	119,260	129,646	119,260
p-value of the difference	.617	.739	.436	.436	.803	.453
Control variables	YES	YES	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In columns 1 and 2, the dependent variable is an indicator variable which is 1 if the student graduated secondary school with Nature/Tech field and 0 otherwise. In columns 3 and 4, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In columns 5 and 6, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. In the odd columns, the cohorts of interest are analyzed. In the even columns, the two cohorts before the cohorts of interest are compared as a test for pretrends. All regressions include control variables on the individual level.

Table A.5: Tertiary education effects: Enrolling in a STEM field

	STEM Degree Enrollment			
	STEM Bachelor		STEM Master	
	Main	Placebo	Main	Placebo
	(1)	(2)	(3)	(4)
Panel A: All students				
Treatment	.005	−.003	.003	.003
	(.004)	(.004)	(.003)	(.003)
Observations	154,042	141,719	154,042	141,719
Panel B: By gender				
Treatment **for women**	.004	−.003	.001	−.001
	(.004)	(.004)	(.003)	(.003)
Observations	81,340	74,890	81,340	74,890
Treatment **for men**	.006	−.004	.006	.007
	(.007)	(.005)	(.005)	(.005)
Observations	72,702	66,829	72,702	66,829
p-value	.803	.901	.317	.243
Panel C: By household income				
Treatment **for low income households**	−.001	−.001	−.006	.001
	(.009)	(.009)	(.006)	(.006)
Observations	30,211	28,530	30,211	28,530
Treatment **for high income households**	.007	−.004	.005	.003
	(.004)	(.004)	(.003)	(.003)
Observations	123,831	113,189	123,831	113,189
p-value	.484	.841	.153	.886
Control variables	YES	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In columns 1 and 2, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM Bachelor and 0 otherwise. In columns 3 and 4, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. In the odd columns, the cohorts of interest are analyzed. In the even columns, the two cohorts before the cohorts of interest are compared. All regressions include control variables on the individual level.

Table A.6: Parents with a STEM degree: Placebo cohorts

	Placebo cohorts		
	Nature/Tech	STEM Bachelor	STEM Master
	(1)	(2)	(3)
Panel A: All students			
Treatment x STEM parents	.013	−.005	−.003
	(.018)	(.014)	(.011)
Treatment	−.002	.001	−.001
	(.004)	(.004)	(.004)
Observations	92,522	92,522	92,522
Panel B1: Women			
Treatment x STEM parents	−.005	−.013	−.0003
	(.015)	(.018)	(.016)
Treatment	−.003	.002	.001
	(.003)	(.006)	(.005)
Observations	48,634	48,634	48,634
Panel B2: Men			
Treatment x STEM parents	.032	.003	−.004
	(.031)	(.020)	(.017)
Treatment	.0001	.0003	−.004
	(.008)	(.006)	(.005)
Observations	43,888	43,888	43,888
Control variables	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field. In column 2, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In column 3, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x STEM parents is a triple interaction term that is 1 if an individual is 'treated' and has at least one parent with a college degree in a STEM field. All regressions include control variables on the individual level.

Table A.7: Parents with a College degree: Placebo cohorts

	Placebo cohorts		
	Nature/Tech	STEM Bachelor	STEM Master
	(1)	(2)	(3)
Panel A: All students			
Treatment x College parents	.001	−.008	−.012
	(.010)	(.009)	(.008)
Treatment	−.003	.003	.002
	(.005)	(.005)	(.004)
Observations	92,522	92,522	92,522
Panel B1: Women			
Treatment x College parents	.007	−.011	−.015
	(.008)	(.012)	(.011)
Treatment	−.006	.003	.006
	(.004)	(.006)	(.005)
Observations	48,634	48,634	48,634
Panel B2: Men			
Treatment x College parents	.009	−.003	−.008
	(.017)	(.013)	(.011)
Treatment	−.001	.001	−.002
	(.009)	(.007)	(.006)
Observations	43,888	43,888	43,888
Control variables	YES	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student graduated secondary school with the Nature/Tech field. In column 2, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Bachelor and 0 otherwise. In column 3, the dependent variable is an indicator variable which is 1 if the student graduated with a STEM Master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x College parents is a triple interaction term that is 1 if an individual is 'treated' and has at least one parent with a college degree. All regressions include control variables on the individual level.

Table A.8: Effect of the policy on STEM major enrollment when having a STEM parent

	Enrollments	
	STEM Bachelor	STEM Master
	(1)	(2)
Panel A: All students		
Treatment x STEM parents	.018	.013
	(.020)	(.014)
Treatment	.0005	−.001
	(.005)	(.003)
Observations	92,522	92,522
Panel B1: Women		
Treatment x STEM parents	.006	.007
	(.022)	(.015)
Treatment	.006	−.001
	(.005)	(.003)
Observations	48,634	48,634
Panel B2: Men		
Treatment x STEM parents	.033	.027
	(.031)	(.022)
Treatment	−.005	−.002
	(.009)	(.006)
Observations	43,888	43,888
Control variables	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM bachelor and 0 otherwise. In column 2, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x STEM parents is a triple interaction term that is 1 if an individual is 'treated' and has at least one parent with a college degree in a STEM field. All regressions include control variables on the individual level.

Table A.9: Effect of the policy on STEM major enrollment when having a College parent

	Enrollments	
	STEM Bachelor	STEM Master
	(1)	(2)
Panel A: All students		
Treatment x College parents	.009	.001
	(.011)	(.007)
Treatment	.0003	−.001
	(.006)	(.004)
Observations	92,522	92,522
Panel B1: Women		
Treatment x College parents	.009	.004
	(.011)	(.008)
Treatment	.005	−.001
	(.006)	(.004)
Observations	48,634	48,634
Panel B2: Men		
Treatment x College parents	.006	−.002
	(.019)	(.013)
Treatment	−.003	.0002
	(.010)	(.007)
Observations	43,888	43,888
Control variables	YES	YES

Notes: * denotes significance at the 10% level, ** denotes significance at the 5% level, and *** denotes significance at the 1% level. Standard errors are displayed in brackets and are robust. In column 1, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM bachelor and 0 otherwise. In column 2, the dependent variable is an indicator variable which is 1 if the student enrolled in a STEM master and 0 otherwise. Treatment shows the point estimate of β_1, the DID estimator. Treatment x College parents is a triple interaction term that is 1 if an individual is 'treated' and has at least one parent with a college degree. All regressions include control variables on the individual level.

B Appendix to Chapter 2

B.1 Institutional setting

This section introduces the Dutch health insurance market. Before 2006, people were insured without deductible either from a collective fund or on the private market. On January 1, 2006, the "Zorgverzekeringswet" (Health Insurance Act) went into effect. This law forced all individuals of age 18 and older to be insured with a private insurance company.

These insurance companies have to offer basic health care with a mandatory deductible of 385 euro (2023). Insurance companies have to offer an additional voluntary deductible of up to 500 euro in exchange for a discount on their premium. At some insurance companies, consumers also opt for voluntary deductibles of 100, 200, 300 or 400 euro, but for simplicity we will focus on the choice between the two most extreme options: the lowest deductible (385 euro) and the highest deductible (885 euro).

As of 2023, there are ten private insurance companies owning 37 brands offering the mandatory health insurance. The prices for the basic insurance for the largest brand per insurance company are displayed in Table B.1. The terms and conditions can differ, but all insurances cover the minimum requirements by the Dutch government. All firms have to publish their prices by a fixed date and are not allowed to modify their prices later.

The average monthly price of the basic insurance without voluntary deductible was 134.23 euro in 2023, while the average monthly price of the same insurance with the highest deductible equaled 114.88 euro. The average yearly discount is therefore 230 euro for an extra risk of 500 euro. This discount leads to a net maximal loss of on average 270 euro per year when one chooses the high deductible.

This means there is a trade-off where the ex-post profit from taking the voluntary deductible in a given year depends on the aggregate health care costs over that year. In Table B.2 the payoffs are shown for opting for a high deductible. With health costs below 630 euros, the high deductible is ex-post profitable. If you had health costs above 630 euros, it would have been profitable to buy the insurance with the low deductible.

Of course, there are many reasons why buyers would prefer to buy the low deductible ex-ante. Especially if expected costs are around the cut-off. First of all, health costs are mostly not perfectly predictable. This uncertainty combined with risk-preferences of a consumer might maximize their utility by choosing the insurance with low deductible. Also, the deductibles are yearly, while the extra premium for the low deductible is paid monthly.

Table B.1: Premiums per month in Euros in 2023 by deductible level

Insurance Brand	385 Euros	885 Euros	Discount
Aevitae (EUcare)	127.95	105.45	22.50
a.s.r. (ASR)	137.50	117.50	20.00
CZ	131.90	114.40	17.50
DSW	137.50	120.00	17.50
Menzis	134.50	114.50	20.00
ONVZ	142.50	122.50	20.00
Salland (ENO)	134.90	109.90	25.00
VGZ	132.95	117.95	15.00
Zilveren Kruis (Achmea)	131.95	116.95	15.00
Zorg en Zekerheid	130.65	112.65	18.00
Average	**134.23**	**114.88**	**19.35**

Source: the websites of the insurance companies consulted on December 15, 2022

Table B.2: Profit from buying insurance with voluntary deductible

Annual health costs (HC) excl GP visits	Profit from choosing high deductible
$HC \leq 385$	230
$385 < HC < 885$	615 - HC
$HC \geq 885$	-270

People who prefer regular costs might desire to pay a higher monthly fee. Finally, lower deductibles lead to weakly lower health care costs for the buyer. Some people might want to prevent the dilemma of whether or not to ask for health care.

All these possible reasons make it hard to claim that it is (partly) due to a status-quo bias that only 12 to 13 percent of the buyers had the voluntary deductible in recent years (van Hijum, 2021), while it would be profitable for almost half of the buyers (van Winssen et al., 2015). A laboratory experiment can get rid of all these confounding factors.

People may also choose to buy voluntary additional insurance packages as for example dental care or physiotherapy on top of their basic insurance, but we will not look into detail into those insurances. Visits to a GP are always completely insured to keep first step of health care accessible for everyone. The consequences like a redirection to the hospital or prescription drugs are to be paid by the deductible.

B.2 Non-standard decision making

The choice of deductible is one that should in theory be re-evaluated every year. However, there are some signs that indicate that the choice might not be made perfectly rational. Besides the standard reluctance to actively think about changing the (deductible of the) health insurance, there are some institutional signs that could lead to non-standard yearly evaluation of the ex-ante optimal deductible.

The first and foremost sign of a bias is that the mandatory deductible is considered to be the standard. All ten aforementioned insurance companies show the prices first with the mandatory deductible of 385 euro and only afterwards you can look at the discount for a voluntary deductible. This means there is already a status-quo where a prospective insurance buyer gets endowed with a low deductible. From there it will evaluate to compare a discount (gain) with an extra downside risk (loss).

If we follow the prospect theory from Kahneman and Tversky (1979) the potential losses of 500 euro could outweigh the gains from the discount from the perspective of the status-quo (the low deductible). This would lead to more buyers choosing the low deductible compared to a scenario where both options would be displayed as two equals.

Another anomaly if you compare this decision making with standard rational decision making, it is notable that there are quite substantial price differences in Table B.1. The switching rate of consumers however is fairly low, only 6.2 percent changed health insurance company in 2019 Dagblad (2020). This is in line with the switching cost literature (e.g. Strombom et al. (2002), Klemperer (1995)). This would go into the direction of Ritov and Baron (1992) which explains the reluctance of consumers to change the status-quo. If consumers do not act they keep the same deductible besides the same insurance company as the year before.

Of course, all the simultaneous changes in 2006 make it impossible to disentangle a single bias. Therefore, the experiment can be used to judge only the status-quo. We introduce a single status-quo to see the effect of a bias in a treatment group on long run equilibrium deviations, keeping everything else equal.

B.3 Experimental Instructions

General Instructions for the control group

Welcome and thank you for participating. In this experiment, you will have the opportunity to earn a certain number of "ECUs" (experimental currency units). At the end of today's session, we will convert these ECUs to an amount of money, using an exchange rate of 10 ECUs = 1 Euro. We will pay you via PayPal. If you do not have PayPal, we will be in contact with you about the transfer method.

Because you showed up on time, you have received a show up fee of 50 ECUs. How much money you earn on top of that will depend on your decisions - so please follow these instructions carefully. *You have 5 minutes to read through 4 pages of instructions.*

Please note the following before we begin: It is very important to us that all participants are focused exclusively on their own decision making. Please do not communicate with the other participants by any means. If you have a question during the experiment, just contact the experimenter in the Zoom-room.

What is this about?

Today's experiment will simulate an insurance market with **6 buyers** and **4 sellers**. Some people in this group will assume the role of sellers and others will be buyers. Who will assume which role will be determined randomly by the computer at the beginning of the session. Once the roles have been allocated, your role will remain the same for the entire session. Because the buyer role and the seller role are different and involve different actions, you will receive specific additional instructions according to your role on screen before your first decisions. First, however, the basic setting is explained in the following.

The session today will consist of 21 periods. After period 9 there will be a minor change, which will be announced on screen in due time. After the 21 periods have been completed, ONE period out of periods 4-21 will be randomly drawn to determine your payoff (you can think of periods 1-3 as practice periods). Both buyers and sellers start every period with a certain amount of ECUs. How your payoff is calculated exactly depends on whether you are a buyer or a seller and on the decisions you make, so this will be explained on screen later.

The insurance market

The market in which you will operate today is a **market for insurance**. As a buyer, you are in each period confronted with a certain probability that you suffer a *damage* (of 200 ECUs). This probability (the **"personal risk"**) differs from buyer to buyer but will remain the same for you in every period. The computer will inform you about your personal risk before you make any decisions. Note that only you will know your personal risk. The sellers will not be able to see it. They will only receive information on the average and on the distribution of the buyers' personal risk.

Every buyer **must** buy insurance against their personal risk. However, the insurance product comes with a **deductible**. A **deductible** is the amount of money that buyers must contribute themselves, in case they suffer a damage (this means the insurance is not perfect). Two different products are available to buy: Product A and product B. The products on offer are identical, with the exception of the **magnitude** of the deductible.

As you will see, the deductible for product B is larger than that for product A. The deductibles for product A and product B have been predetermined; they will not be chosen by the sellers.

As a seller, you sell both product A and product B, and your task in each period will be to **choose a price for each product**. All sellers choose their prices individually and independently.

As soon as all sellers have chosen their prices, buyers choose which product to buy (either A or B) and from whom to buy (which seller). To be more precise, the buyers are allowed to choose a seller in every period but they can choose a product only *every third period* (in periods 1, 4, 7 and so on). In the other periods, the buyer will have to buy the same product as they bought the period before.

Procedures

To prevent that the experiment takes too long, every screen (including this one) has a timer in the upper right corner. The experiment continues automatically if the timer expires.

If a buyer does **not** buy a product from a seller before the timer expires, the buyer's payoff in that period will always be 0 ECUs. This is regardless of whether the buyer has got a damage in that period. It is important to note that the potential payoffs with insurance are never lower than 0 ECUs, even if you suffer a damage with the most expensive insurance.

If a seller does not submit their price before the timer expires, the seller automatically charges a random price for the product.

The experiment will continue automatically when the timer on this screen expires. On that screen, you will receive your role for this session.

General Instructions for the treatment group

Welcome and thank you for participating. In this experiment, you will have the opportunity to earn a certain number of "ECUs" (experimental currency units). At the end of today's session, we will convert these ECUs to an amount of money, using an exchange rate of 10 ECUs = 1 Euro. We will pay you via PayPal. If you do not have PayPal, we will be in contact with you about the transfer method.

Because you showed up on time, you have received a show up fee of 50 ECUs. How much money you earn on top of that will depend on your decisions - so please follow these instructions carefully. *You have 5 minutes to read through 4 pages of instructions.*

Please note the following before we begin: It is very important to us that all participants are focused exclusively on their own decision making. Please do not communicate with the other participants by any means. If you have a question during the experiment, just contact the experimenter in the Zoom-room.

What is this about?

Today's experiment will simulate an insurance market with **6 buyers** and **4 sellers**. Some people in this group will assume the role of sellers and others will be buyers. Who will assume which role will be determined randomly by the computer at the beginning of the session. Once the roles have been allocated, your role will remain the same for the entire session. Because the buyer role and the seller role are different and involve different actions, you will receive specific additional instructions according to your role on screen before your first decisions. First, however, the basic setting is explained in the following.

The session today will consist of 21 periods. After period 9 there will be a minor change, which will be announced on screen in due time. After the 21 periods have been completed, ONE period out of periods 4-21 will be randomly drawn to determine your payoff (you can think of periods 1-3 as practice periods). Both buyers and sellers start every period with a certain amount of ECUs. How your payoff is calculated exactly depends on whether you are a buyer or a seller and on the decisions you make, so this will be explained on screen later.

The insurance market

The market in which you will operate today is a **market for insurance**. As a buyer, you are in each period confronted with a certain probability that you suffer a *damage* (of 200 ECUs). This probability (the **"personal risk"**) differs from buyer to buyer but will remain the same for you in every period. The computer will inform you about your personal risk before you make any decisions. Note that only you will know your personal risk. The sellers will not be able to see it. They will only receive information on the average and on the distribution of the buyers' personal risk.

Every buyer **must** buy insurance against their personal risk. However, the insurance product comes with a **deductible**. A **deductible** is the amount of money that buyers must contribute themselves, in case they suffer a damage (this means the insurance is not perfect).

The deductible for the insurance has been predetermined; they will not be chosen by the sellers.

As a seller, you sell the insurance product, and your task in each period will be to **choose a price for the product**. All sellers choose their prices individually and independently.

As soon as all sellers have chosen their prices, buyers choose from whom to buy (which seller).

Procedures

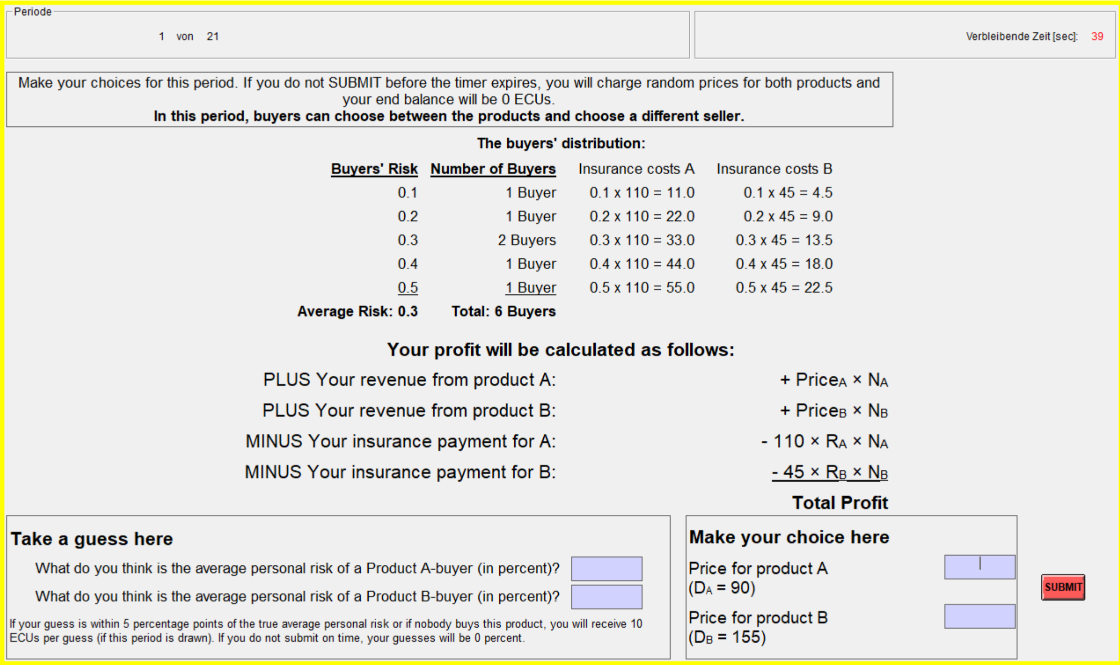
To prevent that the experiment takes too long, every screen (including this one) has a timer in the upper right corner. The experiment continues automatically if the timer expires.

If a buyer does **not** buy a product from a seller before the timer expires, the buyer's payoff in that period will always be 0 ECUs. This is regardless of whether the buyer has got a damage in that period. It is important to note that the potential payoffs with insurance are never lower than 0 ECUs, even if you suffer a damage with the most expensive insurance.

If a seller does not submit their price before the timer expires, the seller automatically charges a random price for the product.

The experiment will continue automatically when the timer on this screen expires. On that screen, you will receive your role for this session.

Figure B.1: Seller's pricing screen



Buyer's RV screen in the control group

You are a buyer. The probability that you suffer a damage is θ percent. Your task now is to choose a **product type (A or B)**.

As you can see from the table, your own contribution in case of a damage (the deductible) will be lower if you choose product A (90 ECUs) than if you choose product B (155 ECUs). On the other hand, the price of product A may be higher than the price of product B. If the price difference between A and B is very small, it makes sense to buy product A. If the price difference is very high, it makes sense to buy product B. The question is:

At what price difference are you indifferent between A and B?

Think carefully before you submit a number. After you submitted your number, the computer will check the **actual prices that are on offer for product A and product B in this period**. Specifically, the computer will compare the lowest available price for A and the lowest available price for B. And then the following rules apply:

(1) If the actual price difference turns out to be **lower** than the number you submitted, you purchase product A.

(2) If the actual price difference turns out to be **higher** than the number you submitted, you purchase product B.

(3) If the actual price difference turns out to be **equal**, the computer chooses randomly which product you purchase.

(4) If you fail to submit a number before the timer expires, you will earn 0 ECUs in the next three periods. The next time you can choose between product A and product B will be in three periods.

Please enter a value between 0 and 65 ECUs and press OK to confirm.

Enter the price difference here:

Buyer's RV screen in the treatment group: product Low

You are a buyer. The probability that you suffer a damage is θ percent. Your current insurance product is 'product A' If you wish, you can now switch to product B for the next three periods.

As you can see from the table, your own contribution in case of a damage (the deductible) will be **higher** if you switch to product B (155 ECUs) than if you stay with product A (90 ECUs). On the other hand, the price of product B may be **lower** than the price of product A. The question is:

What price reduction would be required to make you want to switch to product B?

Think carefully before you submit a number. After you submitted your number, the computer will check the **actual prices that are on offer for product A and product B in THIS period**. Specifically, the computer will compare the lowest available price for A and the lowest available price for B. And then the following rules apply:

(1) If the actual price reduction for switching to product B turns out to be **lower** than the number you submitted, you will continue to get product A.

(2) If the actual price reduction for switching to product B turns out to be **higher** than the number you submitted, you will switch to product B.

(3) If the actual price reduction for switching to product B turns out to be **equal**, the computer chooses randomly whether or not you switch.

(4) If you fail to submit a number before the timer expires, you will earn 0 ECUs in the next three periods. The next time you can switch products will be in three periods.

Please enter a price reduction between 0 and 65 ECUs and press OK to confirm.

Enter the price reduction you require for a switch here:

Buyer's RV screen in the treatment group: product High

You are a buyer. The probability that you suffer a damage is θ percent. Your current insurance product is 'product B' If you wish, you can now switch to product A for the next three periods.

As you can see from the table, your own contribution in case of a damage (the deductible) will be **lower** if you switch to product A (90 ECUs) than if you stay with product B (155 ECUs). On the other hand, the price of product A may be **higher** than the price of product B. The question is:

What price increase would you be willing to accept to switch to product A?

Think carefully before you submit a number. After you submitted your number, the computer will check the **actual prices that are on offer for product A and product B in THIS period**. Specifically, the computer will compare the lowest available price for A and the lowest available price for B. And then the following rules apply:

(1) If the actual price reduction for switching to product A turns out to be **higher** than the number you submitted, you will continue to get product B.

(2) If the actual price reduction for switching to product A turns out to be **lower** than the number you submitted, you will switch to product A.

(3) If the actual price reduction for switching to product A turns out to be **equal**, the computer chooses randomly whether or not you switch.

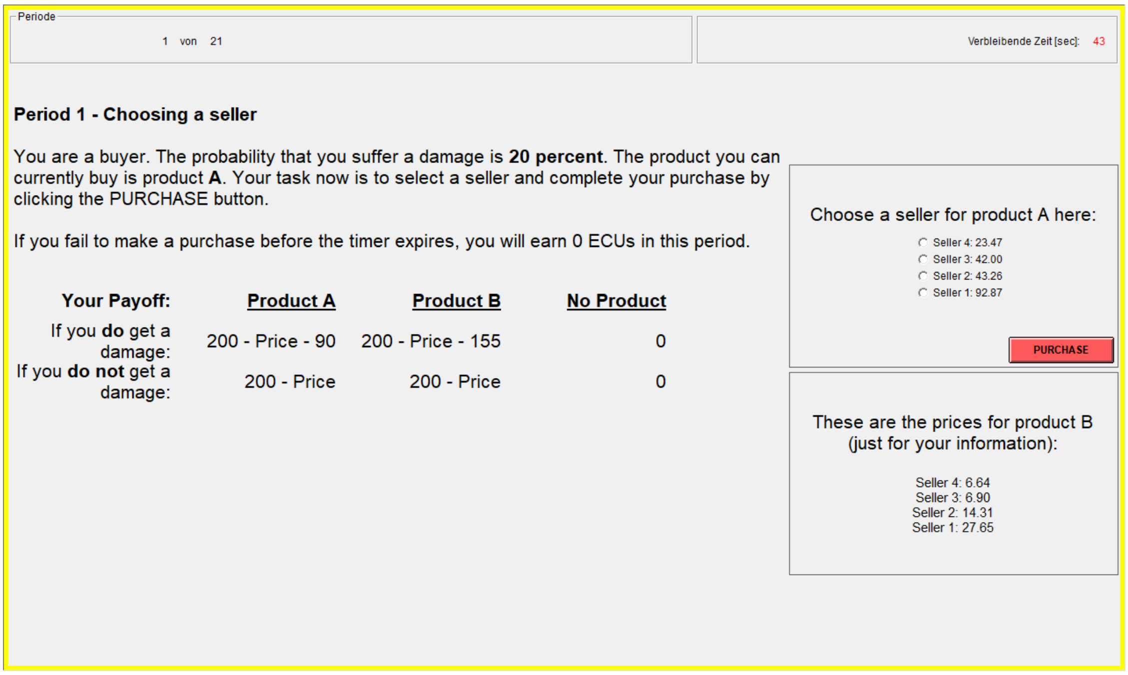
(4) If you fail to submit a number before the timer expires, you will earn 0 ECUs in the next three periods. The next time you can switch products will be in three periods.

Please enter a price reduction between 0 and 65 ECUs and press OK to confirm.

Enter the price increase you are willing to accept here:

Buyer's price selection screen

Figure B.2: Buyer's price selection screen



B.4 Cognitive reflection problems

1. A bat and a ball cost EUR1.10 in total. The bat costs EUR1.00 more than the ball. How much does the ball cost? (in cents)

2. If it takes 5 minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets (in minutes)?

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake (in days)?

4. If three elves can wrap three toys in hour, how many elves are needed to wrap six toys in 2 hours?

5. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?

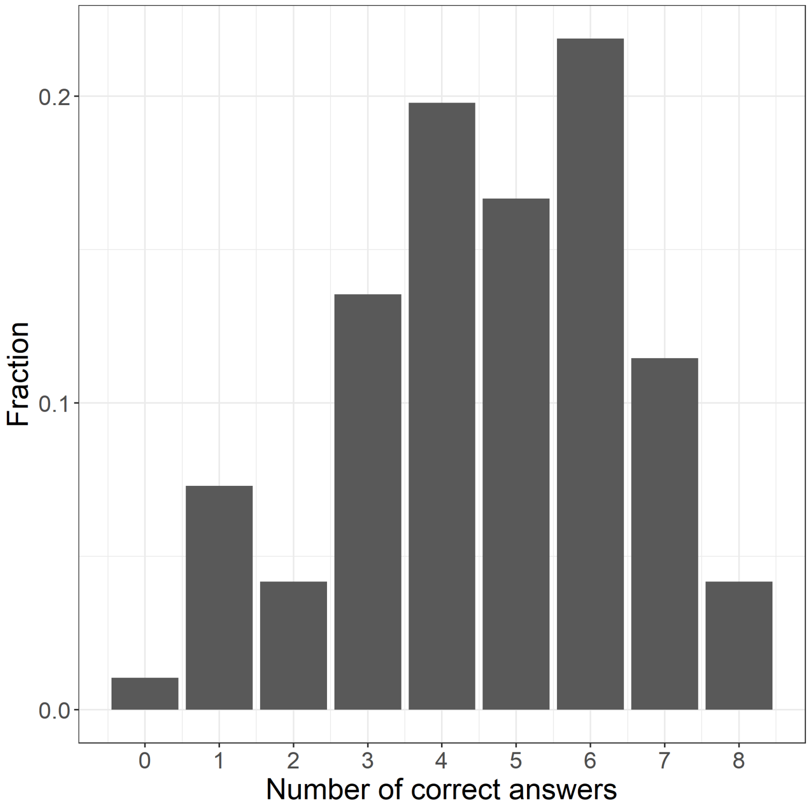
6. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

7. A car and a bus are on a collision course, driving toward each other. The car is going 70 kilometers an hour. The bus is going 80 kilometers an hour. How far apart are they one minute before they collide (in meters)?

8. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together (in days)?

sources: Primi et al. (2015) and Frederick (2005)

Figure B.3: Distribution of CRT scores



B.5 Risk aversion dilemmas

Choose between one of the two following options:
Option X pays out 115.5 ECUs with probability p and 3 ECUs with probability $1 - p$.
Option Y pays out 60 ECUs with probability p and 48 ECUs with probability $1 - p$.

1. p = 0.1
2. p = 0.2
3. p = 0.3
4. p = 0.4
5. p = 0.5
6. p = 0.6
7. p = 0.7
8. p = 0.8
9. p = 0.9
10. p = 1.0

source: Holt and Laury (2002)

The risk aversion dilemmas were presented to the buyers in a random order. Due to this setup it was more likely that subjects made inconsistent choices. To deal with this, the buyers' true risk aversion switching point (RASP) was estimated with a logistic model like in Engel and Kirchkamp (2019).

B.6 Loss aversion dilemmas

Assume that a coin is thrown. The coin can either land at "heads" or "tail". To answer each of the ten questions you will either have to chose "accept" or "reject".

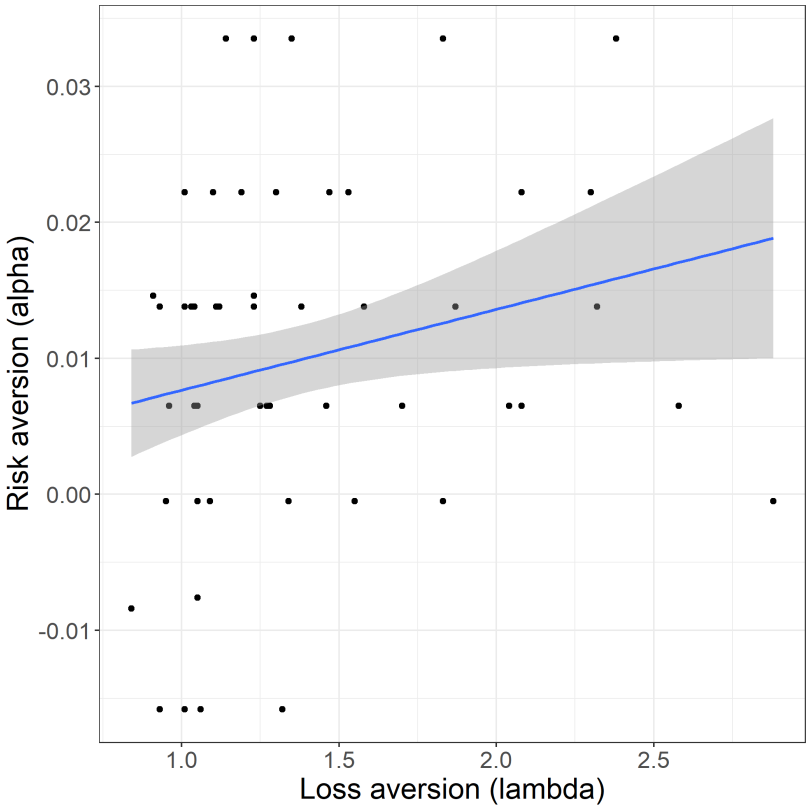
If the coin shows "heads" you will lose x ECUs; if it shows "tail" you will win 50 ECUs. accept / reject?

1. x = 10
2. x = 15
3. x = 20
4. x = 25
5. x = 30
6. x = 35
7. x = 40
8. x = 45
9. x = 50
10. x = 55

source: Rau (2015)

The risk aversion dilemmas were presented to the buyers in a random order. Due to this setup it was more likely that subjects made inconsistent choices. To deal with this, the buyers' true loss aversion switching point (LASP) was estimated with a logistic model like in Engel and Kirchkamp (2019).

Figure B.4: Relation between risk aversion and loss aversion



B.7 Proofs

Proof that there is at most one indifferent consumer

Consumer θ^* is indifferent if
$U(L, \theta^*, r) - U(H, \theta^*, r) = 0$ or
$\theta^* * u(E - p_l - d_l, r) + (1 - \theta^*) * u(E - p_l, r) - \theta^* * u(E - p_h - d_h, r) - (1 - \theta^*) * u(E - p_h, r) = 0$

To show that there is at most one indifferent consumer, it suffices to show that $U(L, \theta, r) - U(H, \theta, r)$ is strictly increasing in θ or $\frac{\partial}{\partial \theta}(U(L, \theta, r) - U(H, \theta, r)) > 0 \forall \theta$.

$\frac{\partial}{\partial \theta}(U(L, \theta, r) - U(H, \theta, r))$
$= u(E - p_l - d_l, r) - u(E - p_l, r) - u(E - p_h - d_h, r) + u(E - p_h, r)$
$= u(E - (p_l + d_l), r) - u(E - (p_h + d_h), r) + u(E - p_h, r) - u(E - p_l, r)$
$u(E - (p_l + d_l), r) - u(E - (p_h + d_h), r) > 0$ given $p_l + d_l < p_h + d_h$ as $p_l - p_h < d_h - d_l$ and
$u(E - p_h, r) - u(E - p_l, r) > 0$ given $p_h < p_l$ so
$u(E - (p_l + d_l), r) - u(E - (p_h + d_h), r) + u(E - p_h, r) - u(E - p_l, r) > 0 \square$

Proof of theorem 2

Firms can only get zero (expected) profits in equilibrium

Proof by contradiction: Assume wlog $M = 2$.

Due to the fact that firms can always price anything above average marginal costs: $\pi_1 \geq 0$ and $\pi_2 \geq 0 \rightarrow \pi_1 + \pi_2 \geq 0$. Suppose $\pi_1 + \pi_2 > 0$, ϵ small enough and wlog $\pi_1 \geq \pi_2$.

If (1) $p_2 > p_1 > c$, $\pi_2 = 0$, but if $p_2 = p_1 - \epsilon$, $\pi_2 = (p_1 - \epsilon - c)D(p_1 - \epsilon) > 0$ and $\pi_1 = 0$. A contradiction.

If (2) $p_2 = p_1 > c$, $\pi_2 = 0.5 * (p_2 - c)$, but if $p_2 = p_2 - \epsilon$, $\pi_2 = (p_2 - \epsilon - c)D(p_1 - \epsilon) > 0.5 * (p_2 - c)$ and $pi_1 = 0$. A contradiction.

Firms can only get zero (expected) profits for each of the packages $\{L, H\}$

Proof by contradiction: Assume wlog $M = 2$.

Suppose $\pi_1 = 0$, but $\pi_1^H > 0$, then $\pi_1^L < 0$. If you reprice $p_1^L > c_l$ then $\pi_1^L \geq 0$ and $\pi_1 > 0$. A contradiction.

Proof that the indifferent consumer has lower risk with an additive bias

The first order derivative with respect to γ is shown below, for all $r \neq 1$. This derivative is always negative. Given that $(E - p_l) < (E - p_h)$ (as $p_l > p_h$) and $(E - p_h - d_h) < (E - p_l - d_l)$ (as $p_l - p_h < d_h - d_l$), we know that the denominator is negative if $r < 1$ and positive if $r > 1$. The numerator, of course, is positive if $r < 1$ and negative if $r > 1$. This implies $\theta^*(\gamma) \leq \theta^*(0) \forall \gamma > 0$.

$$\frac{\partial \theta^*}{\partial \gamma} = \frac{(1-r)}{(E - p_h - d_h)^{1-r} - (E - p_l - d_l)^{1-r} + (E - p_l)^{1-r} - (E - p_h)^{1-r}}$$

Proof of long run equilibrium without bias

We know $r = 0.8$, take the set-up of Table 2.2. Then the only equilibrium is displayed in Table 2.3. Any other set of prices would lead to incentives for at least one buyer or seller to deviate. To determine an equilibrium we have to make sure that (i) none of the sellers want to deviate and (ii) all buyers given the prices chose the product which yields them the highest utility. Given the discrete case of our market we can not merely solve the equations 2.2 and 2.3, but we would have to consider case by case.

Assume $\theta^* > 0.5$, all buyers want the high deductible and $p_h = 0.3 * 45 = 15$. Then $p_l > 64.5$ necessary to keep the preference for the high deductible for the buyer with $\theta = 0.5$. Then a deviation is possible for a seller by setting $p_l \in [58.7, 64.5]$. This would give a positive profit as in this case $\theta^* \in (0.4, 0.5)$. With this θ^*, the prices under Bertrand would go to $p_h = 11.7$ and $p_l = 55$.

This would then give an incentive to deviate to $p_l \in (49.5, 54.5)$. This would lead to $\theta^* \in (0.3, 0.4)$ and positive profits for the deviating seller. The prices under Bertrand would now go to $p_h = 10.125$ and $p_l = 49.5$.

This would lead to a seller deviating to $p_l \in (44, 45.3)$. This would lead to $\theta^* \in (0.2, 0.3)$ and positive profits for the deviating seller. The prices under Bertrand would now go to $p_h = 6.25$ and $p_l = 41.25$. However, this price setting, which would lead to zero profit if $\theta^* \in (0.2, 0.3)$, would also lead to $\theta^* = 0.3$ with $U(L, 0.3) \approx U(H, 0.3) \approx 13.13876$, this is a contradiction and would result into constant coin flips from θ^*-buyers.

This can never be a sTable equilibrium as in expectation one buyer will buy product $L$and the other will buy product H. $\theta^* = 0.3$ would lead to prices $p_h = 9$ and $p_l = 44$ and $U(L, 0.3) \approx U(H, 0.3) \approx 13.077$. On average, one θ^*-buyer will buy product $L$and one would buy product H. This is the equilibrium from Table 2.3.

If sellers want to deviate again, they have to lower their prices again as this would be the only way to lure buyers away from competitors. However, lowering their prices to get a $\theta^* \in (0.2, 0.3)$ would not work as we would be in the same situation as above. $\theta^* \in (0.1, 0.2)$ would not be an equilibrium as this would lead to $p_l = 37.4$ and $p_h = 4.5$. Then $\theta^* > 0.2$ which is a contradiction.

$\theta^* < 0.1$ would not be an equilibrium as this would lead to $p_l = 33$ and $p_h > 17.6$. This would lead to a seller charging $p_l > 33$ and $p_h \in (4.5, 17.6)$ to persuade the buyer with $\theta = 0.1$ to buy the high deductible. With an expected cost of 4.5 for this consumer, this seller would get a positive profit.

For completeness, $\theta^* \in \{0.1, 0.2, 0.4, 0.5\}$ are all no equilibrium options as the prices the sellers would charge to get zero profits lead in all cases to the consumer with the respective θ^* not to be indifferent. Which would be a contradiction. Q.E.D.

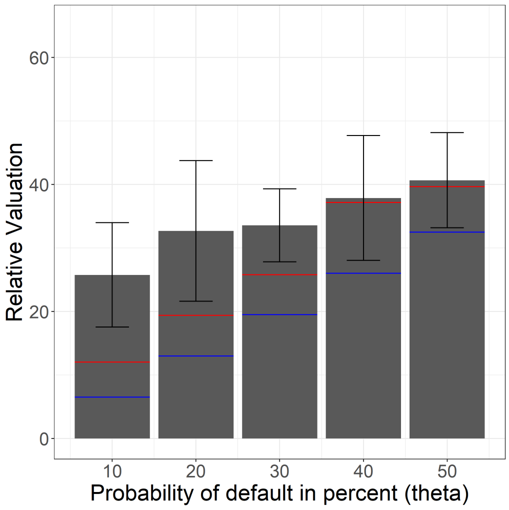
B.8 Sample characteristics

Table B.3: Sample characteristics: Buyers

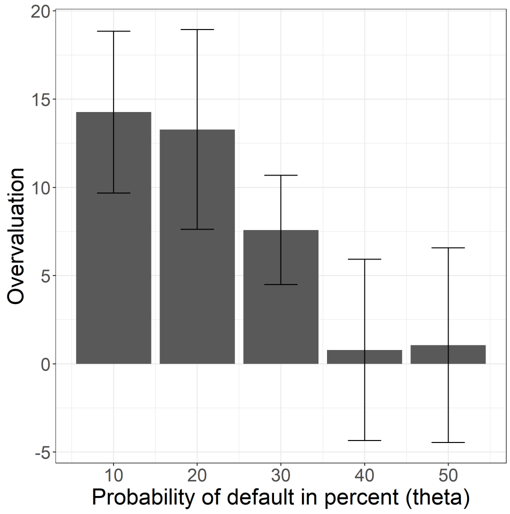
	Treatment Low	Treatment High	Control
N	36	36	24
Female	0.472	0.333	0.583
Age	24.36	22.14	24.04
Economics	0.306	0.222	0.25
German	0.777	0.806	0.833
CRT score	4.888	4.583	4.333
Loss aversion	1.595	1.576	1.777
Risk aversion	0.427	0.418	0.613
Product Low purchase	0.729	0.736	0.726

B.9 Additional Figures

Figure B.5: Adverse selection on the market by risk level



(a) RV for Product Low



(b) Fraction of Product Low-buyers

Note: The y-axis in subfigure (a) is the average reported RV for product Low. The y-axis in subfigure (b) is the average overvaluation for product Low. The blue lines are the RV levels for risk neutral buyers and the red lines show the expected RVs corresponding to the risk-aversion parameter r as measured by the Holt-Laury task. The whiskers indicate 95 percent confidence intervals.

C Appendix to Chapter 3

C.1 Experimental Instructions

Welcome and thank you for participating. In this experiment you will have the opportunity to earn a certain number of "ECUs" (Experimental Currency Units). At the end of today's session, we will convert these ECUs to an amount of money, using an exchange rate of 8 ECUs = 1 Euro. As you showed up on time, you earned a show-up fee of 40 ECUs. This is added to your final amount. How much money you earn on top of your show-up fee will depend on your decisions – so please follow these instructions carefully.

The experiment of today consists of 10 different and independent tasks. These tasks will all be individually explained on screen in due time. All tasks will give you task-specific earnings depending on your decisions.

At the end of the experiment, one of the 10 tasks will be selected by the computer completely at random. Only this task's earnings will be added to your show-up fee. All earnings you make in the other 9 tasks will not be paid out. Therefore, if a task states for example: 'You receive ECU100.00', you will only receive the earnings if this task is drawn by the computer.

We will pay you in cash and in private. When we pay you, we will only see the final amount you have earned and not how much you have earned in each task or which task we pay out.

All pages have a timer to ensure a timely progression. Be aware that if the timer expires, the experiment will save your input and thereafter move on to the next screen. Please ensure that you make all your decisions before the timer expires.

Please note the following before we begin: It is very important to us that all participants are focussed exclusively on their own decision making. Please do not talk to the other people and also do not communicate with them by any other means. If you have a question during the experiment, just raise your hand and we will come to you.

Figure C.1: Instructions round 1

Task 1: Move the Sliders - Round 1: Instructions

Time left to complete this page: **1:46**

This task consists of two rounds. In this task you can score points by moving a slider's thumb into the correct position. The thumb is initially located on the left and this location corresponds to a slider value of 0. The maximum slider value is 100 which corresponds to the thumb being positioned on the right edge.

Your goal is to reposition the thumb such that the slider value becomes 50. To move the thumb, click on it with your mouse and then drag it to the centre of the slider. Whenever you release the thumb, the slider value – shown on the right-hand side – will be updated according to the thumb's current location. You can move the thumb as often as you like. As soon as you have succeeded in setting the slider to the target value of 50 click on the 'Next' button. A new screen will then show up and will give you a new slider. You will have 120 seconds for this round, and your point score is determined by the number of sliders you are able to set to 50 during this time.

You will compete with 3 randomly selected people in this round. If your point score is higher than that of your 3 competitors, you win and your earnings for this round will be ECU8.00 per point scored. If another person scored better than you, you lose and your earnings for this round are ECU0.00. Ín case of a tie between two or more competitors, the winner of this round will be selected at random.

The round begins automatically when the timer on this screen expires. Round 2 will be explained after round 1 ends.

Figure C.2: Slider task

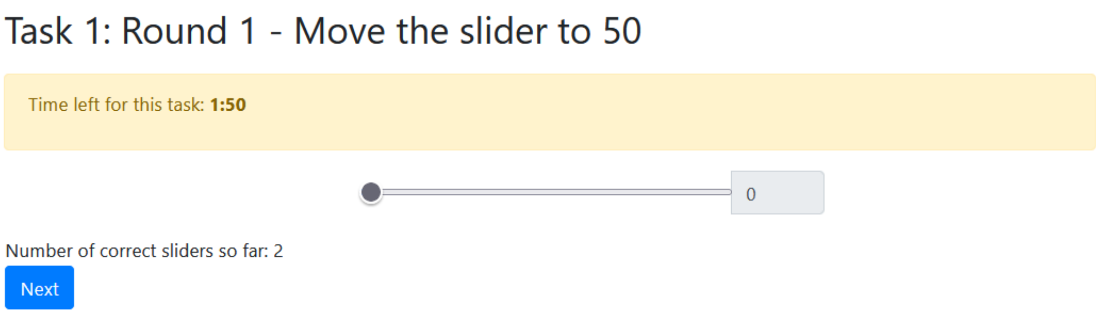


Figure C.3: Results screen round 1

Task 1: Results of round 1

Time left to complete this page: **0:47**

You are player:	**4**
Score of player 1:	0
Score of player 2:	0
Score of player 3:	0
Score of player 4:	27
The winner of this round is player	4
Your earnings for this round:	ECU216.00

Wait for the timer to expire to move to the next page.

Figure C.4: Instructions round 2 and payment scheme choice

Task 1: Move the Sliders - Round 2: Instructions

Time left to complete this page: **1:32**

In this round you can again score points by moving a slider's thumb into the correct position. You will again have two minutes to move as many thumbs to 50 as possible.

You now have the choice between 'Fixed Payment' and 'Competition'.

If you choose 'Fixed Payment', your earnings for this round will be ECU2.00 per point. This is regardless of how many points other participants score.

If you choose 'Competition', you will be matched afresh with 3 randomly selected people in this round. These are not necessarily the same people as you competed against in the first round. In fact, this would be very unlikely. We will compare your score from this round with the scores obtained by these three people in the first round. This means you can be matched with both people who choose the 'Competition' or the 'Fixed Payment' scheme. If your point score is higher than that of your 3 competitors, you win and your earnings for this round will be ECU8.00 per point scored. If at least one of the three persons scored better in round 1 than you do now, you lose and you will receive nothing. Íf your score is tied as best, whether you win or not will be selected at random.

The computer will randomly choose the earnings of one of the two rounds as final earnings for this task.

Do you wish to participate in a competition or receive a fixed payment?

- (●) Competition
- ( ) Fixed payment

The round begins automatically when the timer on this screen expires.

Figure C.5: Results screen round 2 in competition scheme

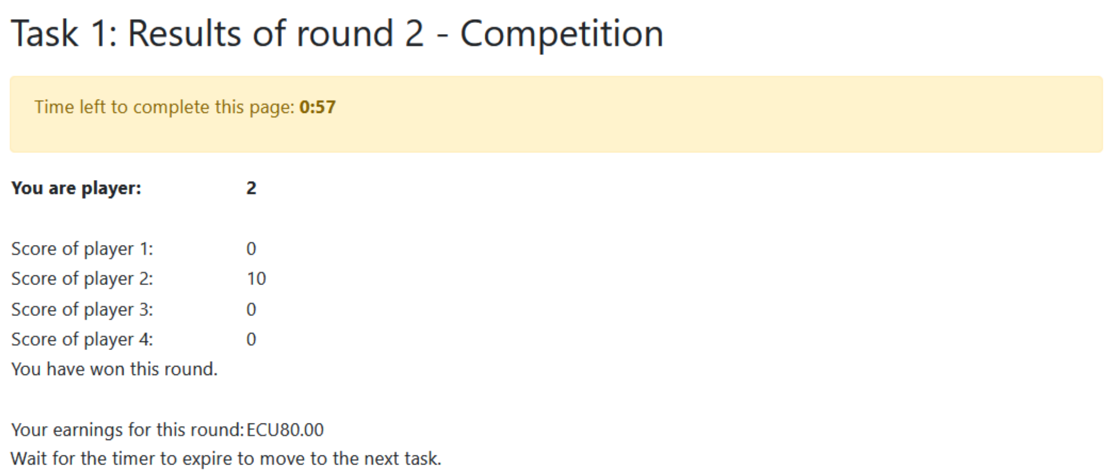
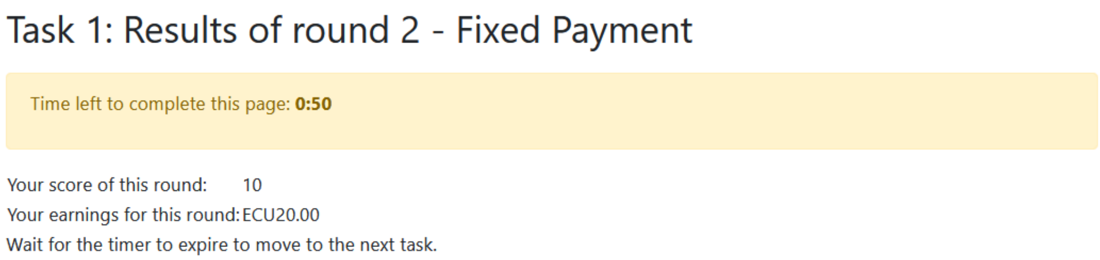


Figure C.6: Results screen round 2 in fixed payment scheme



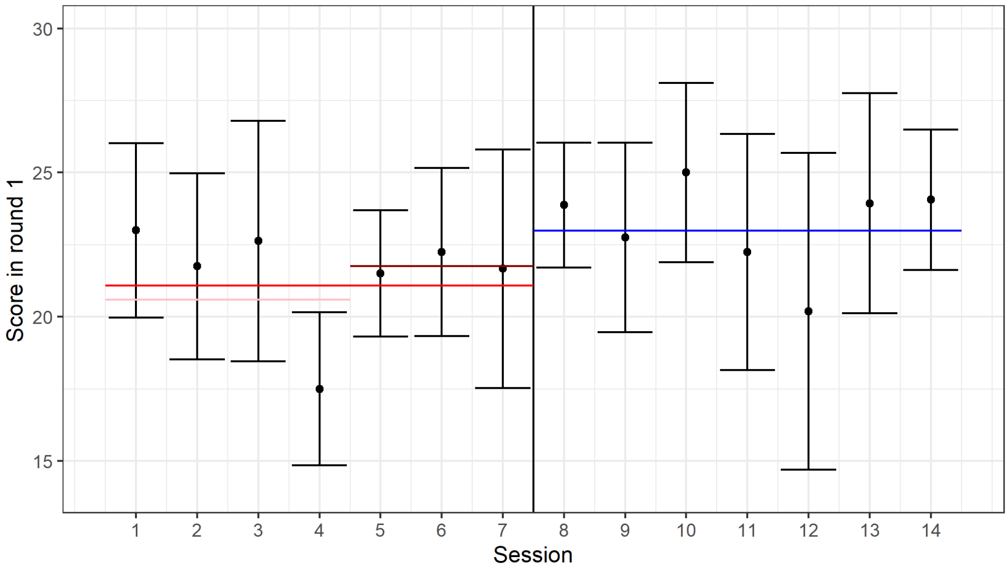
C.2 Questionnnaire

Please indicate the extent to which you agree with the following statements.

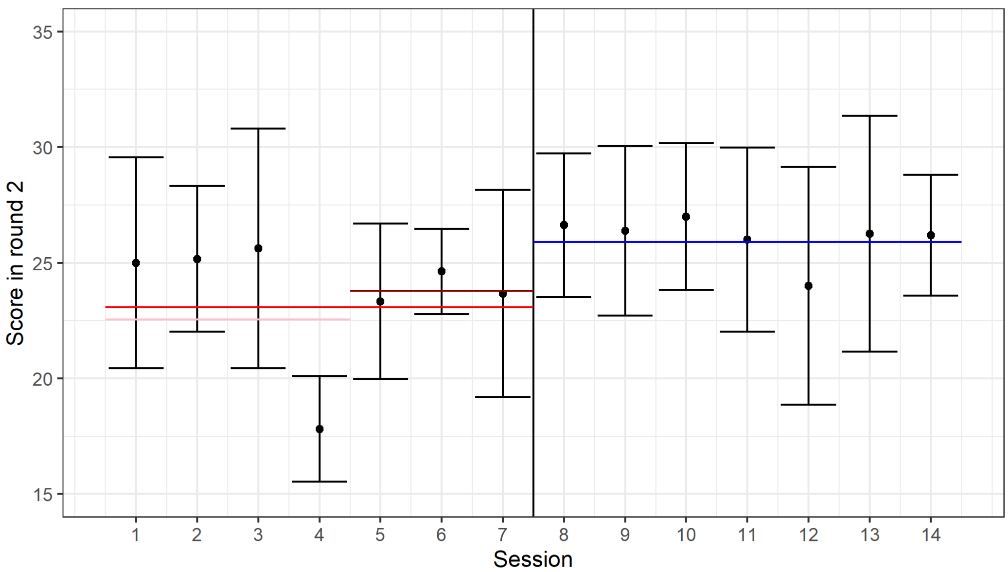
1. I feel well now.
2. I feel warm now.
3. I feel comfortable now.
4. I feel tired now.
5. I feel angry now.
6. I feel bored now.
7. I can breath easily now.
8. I feel satisfied with my group in the tasks
9. I enjoyed participating in this experiment.
10. I think mouth and nose protection are effective in combating the COVID-19 pandemic.
11. I am in favour of mandatory mouth and nose protection in public spaces during the COVID-19 pandemic.

C.3 Additional Figures

Figure C.7: Averages of slider scores by session including session 4



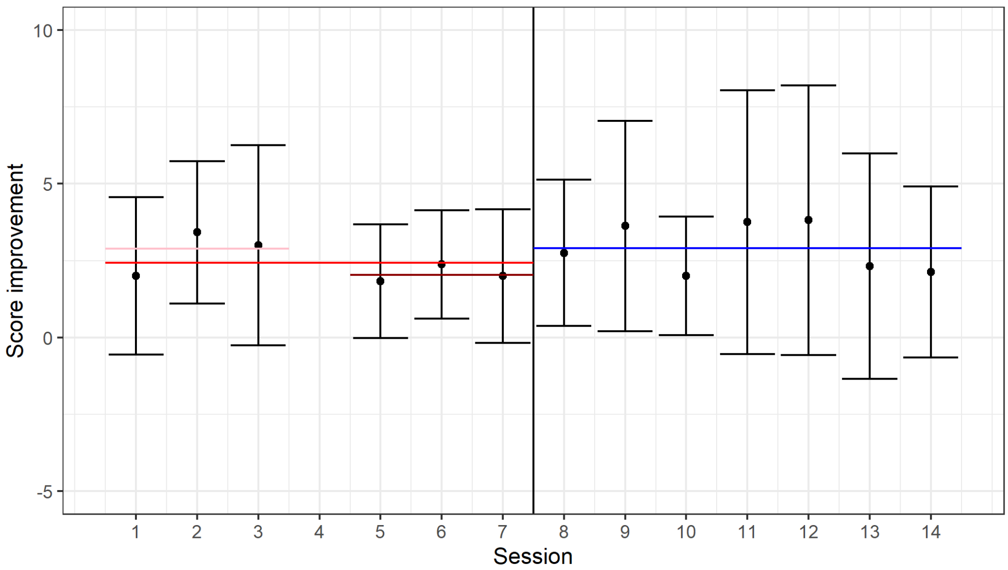
(a) Slider score averages in round 1



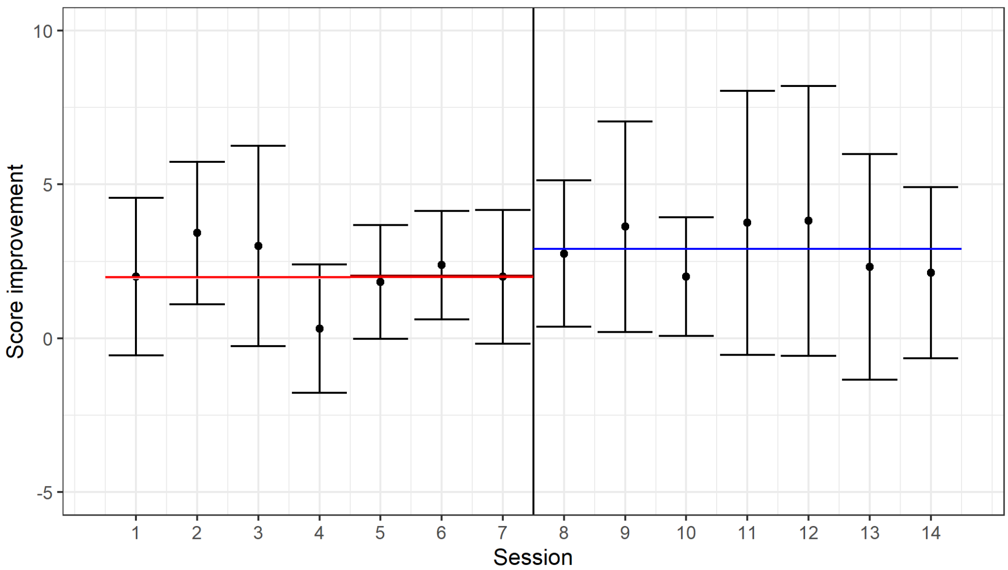
(b) Slider score averages in round 2

Note: The y-axis in subfigures are the average number of correctly calibrated sliders. The blue lines show the averages of all individual scores without masks and the red lines show the averages of all individual scores with masks. The dark red lines are the averages across sessions with FFP2 masks and the lightred lines are the averages across all surgical/FFP2 mask sessions. The whiskers indicate the sessionwide standard deviations.

Figure C.8: Average improvements of slider scores by session



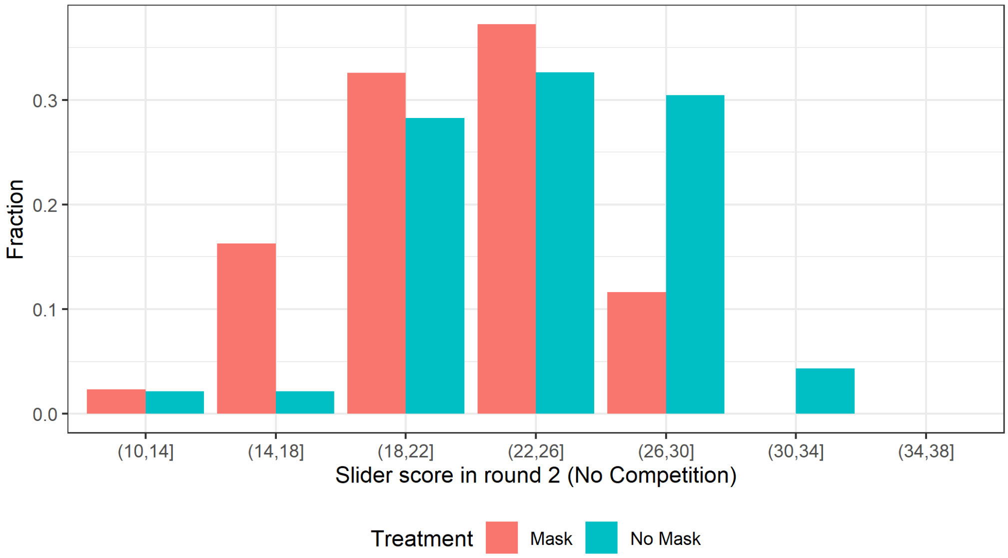
(a) Slider score improvements excluding session 4



(b) Slider score improvements including session 4

Note: The y-axis in subfigures are the average improvements of correctly calibrated sliders in the second round. The blue lines show the averages of all individual scores without masks and the red lines show the averages of all individual scores with masks. The dark red lines are the averages across sessions with FFP2 masks and the lightred lines are the averages across all surgical/FFP2 mask sessions. The whiskers indicate the sessionwide standard deviations.

Figure C.9: Distribution of slider scores on the fixed payments-scheme



[This page intentionally left blank]

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben habe.

Mannheim, 13.02.2023

Mark Jeffrey Spils

[This page intentionally left blank]

Curriculum Vitae

2018-2023	Ph.D. Studies in Economics, University of Mannheim
2016-2018	M.Sc. in Economics, University of Mannheim
2012-2016	B.Sc. in Economics and Business Economics, Vrije Universiteit Amsterdam