Essays in Applied Microeconomics

INAUGURALDISSERTATION

zur Erlangung des akademischen Grades

Doctor Rerum Politicarum (Dr. rer. pol.)

eingereicht an der Wirtschaftswissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

von

Alisa Frey, M.Sc.

Mai 2024

Erstgutachter: Univ-Prof. Dr. Justus Haucap

Zweitgutachter: Univ-Prof. Dr. Alexander Rasch

For all those I love, here or forever in my heart.

Acknowledgements

The completion of this dissertation is due to the invaluable support and feedback of numerous people, not all of whom can be named individually here. Even if this section remains incomplete, I would like to express my deepest gratitude to all those who have shaped and contributed to this work in various ways.

First, I would like to thank my first supervisor Justus Haucap for his various support, not only during the challenging time of writing this dissertation, but also before I officially joined DICE as a Ph.D. student. Second, I am also grateful to my second supervisor Alexander Rasch for his support, guidance, and for always having an open door to chat. A separate thank you also to my supervisors as my co-authors as well as Carina and Nicolas Fugger whom I also co-authored my second dissertation chapter with.

A special thank you goes to Tim Cason and the experimental economics group at Purdue University for welcoming and supporting me during my research stay and beyond. The feedback and encouragement symbolically carried me through the last semester of writing this dissertation, and I am very grateful for that. I would also like to use this opportunity to thank Hans-Theo Normann and other DICE experimenters, such as Hannah-Schildberg-Hörisch, for their support and advice in carrying out my experimental projects.

Furthermore, I would like to thank the DICE and the DFG-funded GRK 1974 for the generous support during my Ph.D. studies. The research environment provided was outstanding and should not be taken for granted. I would also like to thank all (current and former) DICE members for creating an exceptional collegial atmosphere, even during a pandemic. In particular, I would like to thank my officemate and friend Alexandra J. Gibbon for the many laughs, chats, ramen dates, and more – I have always enjoyed our shared office and beyond (and your easily accessible econometric tips). I am also grateful to Tobias Felix Werner for many long chats in the office as well as sharing his Python and otree expertise. Not forgetting to mention my fellow Ph.D. students whom I want to thank, I especially thank Sabrina Herzog, David Zeimentz, Leonard Gregor, Jana Gieselmann, and Judith Schwartz for many fun lunch and coffee chats etc. I would also like to thank Andreas Lichter and Katharina Erhardt for their empathy and supportive attitude towards young scholars.

Finally, I would like to express my special thanks to my family and friends who have supported me all along the way in various ways. They have had to exercise a lot of patience and renunciation during this time. I have not always lived up to my own expectations of being a good daughter, sister, or friend during my PhD, especially not in the final year of this dissertation. I thank you for your understanding, loyalty, and for accepting me for who I am. Without your support and perseverance, this dissertation journey would not have been possible, and I am very grateful for all your encouragement. I am truly eternally grateful for all the adventures along the way.

Contents

List of Figures

List of Tables

Introduction

> *If we want our regulators to do better, we have to embrace a simple idea: regulation isn't an obstacle to thriving free markets; it's a vital part of them.*
>
> James Surowiecki

Regulation plays a vital role in many aspects of our lives. It ensures safety, fairness, and helps reduce asymmetries in markets and society. Social regulation focuses on ensuring a desired level of equality, fairness and social justice. Economic regulation can maintain fair competition and protect consumers from exploitation. As diverse as regulation can be, it always seeks to solve problems that would be harmful if left unaddressed. However, it is equally important to constantly question and review (current) regulation. This cumulative dissertation aims to question and scrutinize regulation in various dimensions.

My co-authors and I examine the regulatory implications for specific markets as well as for the broader state and society at the intersection of industrial and public economics. The primary approach of this work is empirical, yet underpinned by a theoretical framework. In three chapters, the focus zooms from evaluating a policy change for a specific market (chapter 1), to analyzing different market regulations for a specific class of goods (chapter 2), to a societal question of different levels of redistribution (chapter 3).

In the *first chapter* of my dissertation, co-authored with Justus Haucap, we investigate the market for menstrual products in Germany. Our study focuses on an empirical analysis of the impact of the value-added tax (VAT) reduction on menstrual products and the extent to which this tax cut is passed on to consumers.

We provide empirical evidence that contributes to the ongoing discourse on VAT pass-through rates. Specifically, we want to evaluate the regulatory change in the VAT rate for menstrual products and assess its effectiveness in terms of a gross price reduction.

Our findings indicate that the VAT reduction is fully passed through to consumers. Interestingly, we observe that the pass-through rate exceeds 100 percent, a phenomenon we attribute to rounding effects in retail pricing. Additionally, our analysis suggests that media attention had an impact on the pass-through rate, highlighting the role of public in economic outcomes. This research offers important insights for policymakers considering similar VAT adjustments.

Stepping back from a specific product group, the *second chapter* of my dissertation focuses on a specific class of goods: experience goods. For markets with experience goods, there exist information asymmetries between market sides which at least suggest that such markets could need some regulation. Whether a regulatory intervention like an entry barrier favors different market outcomes in such markets, is the research question that is addressed in this chapter. In a market design study (with Nicolas Fugger, Carina Fugger, and Alexander Rasch), we use an economic experiment to investigate the effects of occupational licensing as a barrier to entry on market outcomes in a market for experience goods. In particular, we evaluate our results from a consumer perspective. In some markets, occupational licenses are active barriers to entry due to regulation and partial deregulation in recent decades, while in other markets, these barriers have disappeared. Craft services are a typical example. Various stakeholders support further deregulation, including consumers, chambers, and professionals. To gather reliable data on experience goods markets, the experimental method is preferred due to the limited availability of market information. We find mixed results on whether occupational licensing as a barrier to entry is beneficial for consumers. This depends on the nature of the market and the availability of information.

In our experiment, we examine the effects of occupational licensing as a barrier to entry in two different market scenarios. The scenarios differ in the consequences that result from obtaining the license, these consequences are either the ability to provide a higher quality or the provided quality is transparent to buyers.

On the one hand, we find that an entry barrier lowers social welfare if the investment in qualification only enables the ability to provide a higher quality product, but does not lead to transparency about the quality of the good provided. On the other hand, we see tendencies that consumers benefit from the transparency that is made possible by the investment in an occupational license in terms of a higher utility of an average consumer in this markets.

The *third chapter* of my dissertation is single-authored. In this chapter, I zoom out to societal regulation and a question on redistributive justice. I study the effects of redistributive preferences when citizens can vote with their feet and when their income-generating behavior is influenced by moral hazard. My economic experiment is inspired by John Rawls' "veil of ignorance" thought experiment. The primary research question I investigate is what redistribution people choose over time.

Firstly, I analyze decision-making behind the veil of ignorance. Secondly, I investigate whether decisions alter when participants are aware of their own capabilities and receive information regarding the capabilities of others. Using a theoretical model, I hypothesize about the decision-making process both behind the veil of ignorance and after it is lifted.

While the first question has already been studied in economic and philosophical experiments across multiple decision-making scenarios, the combination of repeated decisions without the veil of ignorance remains largely unexplored. The follow-up question is relevant because it transfers the abstract thought experiment to real-world applications, such as choosing a residence based on proximity to borders between different tax authorities, e.g., states or cantons, including considerations for taxation and social benefits.

In six rounds, participants select either a high or low redistribution group. Redistribution occurs within each selected group based on the principle of voting by feet. Income for each participant is earned after the decision is made and is heterogeneous due to variations in ability, effort, and moral hazard.

Each of these chapters evaluates and challenges a regulatory issue. The variety of regulatory issues is reflected in the equally diverse nature of the chapters in this dissertation. Since all the chapters were conceived as independent research papers, this diversity is not surprising – but the identification of a common denominator is. At the time of writing, only the first chapter has been published. Interested readers of chapters 2 and 3 are encouraged to check for updated versions of the papers.

Chapter 1

VAT Pass-Through: The Case of a Large and Permanent Reduction in the Market for Menstrual Hygiene Products

Co-authored with Justus Haucap[1]

[1]This paper is now published in *International Tax and Public Finance* Vol. 31, Feb 2024, pp. 160-202, DOI 10.1007/s10797-023-09813-w.

1.1 Introduction

Governments often use value-added tax (VAT) rate cuts to stimulate demand (post-2008 crisis[2]), improve consumer spending power (COVID-19 pandemic[3]) or to benefit specific groups (e.g., low-income households). More recently, governments have reduced VAT rates on sanitary products in an effort to make the tax system more equitable from a gender perspective. In Germany, following a widely publicized petition to the German parliament, the VAT on feminine hygiene products, tampons and sanitary pads, was reduced from the standard rate to the reduced rate in 2020. Several countries, as well as individual states in the U.S., have adopted similar tax policies favoring menstrual hygiene products in recent years.[4]

Whether or not policymakers achieve their stated goal depends on the extent to which the tax reduction translates into a price reduction. The question of how much of a VAT reduction is passed on to consumers is a source of debate around these VAT rate changes.[5] Previous evidence on this question suggests ambiguous results. For example, Montag et al. (2021) analyze the effect of a general temporary VAT cut on gasoline prices in 2020, finding less than full pass-through, and Dovern et al. (2022) analyze another temporary fuel tax cut in 2022, finding close to full pass-through, both in Germany.

This paper analyzes the pass-through of a permanent VAT reduction on menstrual hygiene products in Germany. We examine how the gross prices of different product types, and in different store types and competitive settings changed within a time period that includes both the official decision to reduce the VAT and the date on which it took effect (the actual event). We find that retailers not only passed through

[2]The United Kingdom implemented a temporary VAT cut in 2008 and 2009 (see e.g. Blundell, 2009; Crossley et al., 2009).

[3]For example, Germany and Spain temporarily reduced VAT rates (see e.g., Fuest et al., 2021, and the references therein.).

[4]For example, Kenya, Canada, India, Malaysia, and Australia have no VAT on menstrual products, as do the US states of Nevada, New York, Florida, Connecticut, and Illinois (see, e.g., Zraick, 2019 for The New York Times; Masterson, 2022 for World Economic Forum). In addition, several European countries, such as Belgium, Cyprus, France, the Netherlands, Spain and the United Kingdom, have changed the VAT rate applied to menstrual hygiene products to a reduced VAT rate (see online the Taxes in Europe Database (TEDB)). Since 2006, it has been legally permissible for EU Member States to apply only a reduced VAT rate to menstrual hygiene products (see, e.g., OJ L 347, 11.12.2006).

[5]Uncertainty about the tax incidence was, for example, a stated reason why a petition launched in 2015 to abolish VAT on menstrual hygiene products in Germany was not considered further by the Petitions Committee of the German parliament (see Petition 58474).

the full VAT reduction but actually reduced prices by more than the VAT change would suggest.

Several aspects make this VAT reduction particularly interesting. First, the change can be considered exogenous in the sense that it was not triggered by market conditions. There were no changes in supply or demand that led policymakers to adopt the policy. The VAT reduction can be seen as a natural experiment caused by an exogenous shock. Second, the VAT cut was substantial and was made permanent. The tax rate was reduced from 19 percent to seven percent. The reduction of 12 percentage points is relatively large compared to other VAT changes which often amounted to very few percentage points. Further, unlike other VAT policies observed in recent years, this rate reduction was made permanent. This is in contrast to many other VAT reductions that have been temporary. Third, the demand for these products can be assumed to be fairly inelastic within the common price ranges, and any price reduction due to the VAT cut is unlikely to increase demand. There is generally no acceptable way to substitute these products.[6] Short-term demand effects are also unlikely: Since the tax has been reduced, ex ante stockpiling does not make any sense, and postponement of demand is also unlikely for biological reasons. All in all, demand effects should be less of a concern in this setting. These characteristics make this VAT reduction an interesting case study.

The pass-through of this reform is examined using weekly price data at the retailer-product level. We exploit sharp price changes around the time of the reform. The data also allow us to examine robustness, for example, by varying the number of retailers and the week in which the reform took effect.

Our results are as follows. We find full and even excess pass-through. We observe excess pass-through in relatively "competitive market segments": Prices were reduced more (i) for the quantitatively more significant product type (sanitary pads), (ii) in the retail segment specializing in sanitary products (drugstores, as opposed to supermarkets), and (iii) for products that were offered by two or more retailers. The excess pass-through is almost entirely explained by the fact that retailers round (down) prices to popular decimals (like 0, 5, and 9) after the VAT reduction, and that this rounding occurs in the aforementioned competitive segments of the market.

[6] See also Cotropia and Rozema (2018) on both the inelasticity and substitutability issues.

We further document that media attention may indeed have contributed to this behavior.

1.2 Related literature and theoretical background

1.2.1 Literature

This paper mainly contributes to the empirical literature on the pass-through of VAT changes. There is a broad literature investigating the effects of VAT changes on consumer prices. However, these studies mostly focus on the VAT change as a fiscal instrument (e.g., Blundell (2009) and Crossley et al. (2009) on the UK temporary VAT cut in 2008 and 2009). Also, in recent studies the VAT decrease is analyzed foremost in terms of the intended stimulus, like the temporary VAT reduction during the pandemic in Germany in 2020 (see Montag et al., 2021; Fuest et al., 2021). Both studying fuel prices, Montag et al. (2021) find lower pass-through rates for the general temporal VAT cut compared to Dovern et al. (2022) who analyze a special fuel tax cut.

There is evidence that prices respond differently to increases and decreases. Doyle Jr and Samphantharak (2008), Carbonnier (2008) and Benzarti et al. (2020) analyze differences in tax incidence within increases and decreases. While Doyle Jr and Samphantharak (2008) find symmetric price responses for short-term interventions, Carbonnier (2008) and Benzarti et al. (2020) identify asymmetries comparing the effects of VAT increases in contrast to VAT reductions. According to Benzarti et al. (2020) prices respond less to a VAT decrease than to a VAT increase. However, Hindriks and Serse (2022) find full pass-through for both a tax decrease and a tax increase without any asymmetries when analyzing VAT reforms of residential electricity prices in Belgium.

The effects of permanent VAT changes have mostly been studied when all consumer products are concerned (see, e.g., Benedek et al., 2020; D'Acunto et al., 2022), but specific markets and products have also been analyzed. For example, the effects of a permanent VAT decrease for services like hairdressing or housing repairs have been analyzed as have expensive products like cars (see Carbonnier, 2007; Kosonen, 2015). Compared to these products and markets, menstrual hygiene products are in a very different price range and, as already mentioned, have a very inelastic demand.

Additionally, this paper contributes to the small literature on menstrual hygiene products and taxation. Cotropia and Rozema (2018) have also studied the effects of eliminating sales taxes on menstrual hygiene products. In this paper, they focus on demand-side differences and the distributional effects of tax policy. They use data on purchases of menstrual hygiene products in New Jersey and surrounding states from 2004 to 2006 and study the incidence of the tax using a difference-in-differences approach. Cotropia and Rozema (2018) find that the tax cut is fully shifted to consumers, but is not equally distributed across different consumer groups. While the tax cut reduced prices for low-income consumers by more than the size of the tax, the tax repeal reduced prices for high-income consumers by less than the size of the tax. They add to these results that the pass-through estimates are partly driven by changes in the use of coupons. Cotropia and Rozema (2018) also attempt to find supply-side differences to explain the pass-through variation for different consumer groups. For pass-through differences by income, they find no significant differences across products in a store but within a product in different stores.

In contrast to Cotropia and Rozema (2018), our paper analyzes supply-side effects and differences. In addition to analyzing different competitive segments, we focus on explaining the observed excess price reduction and provide valuable insights into firm pricing.

Furthermore, Rüll (2020) explores different strategies to reduce the tax burden on menstrual hygiene products. The paper focuses on the possibility of promoting sustainable products and discouraging companies along the supply chain from increasing their profit margin when the tax burden is reduced. For a more recent review of different tax policies and their impact on the affordability of menstrual hygiene products, focusing on low- and middle-income countries, see Rossouw et al. (2020).

Our paper also provides some insights into the relationship between consumer attention and retail pricing. As has been shown in several papers (see, e.g., Dickson and Sawyer, 1990; Evanschitzky et al., 2004; Vanhuele and Drèze, 2002; Vanhuele et al., 2006; Loy et al., 2020), consumers' knowledge and memory of retailer prices is highly imperfect. It appears that price knowledge is more or less limited to a relatively small range of products that are consumed regularly. As a consequence, retailers tend to compete most vigorously on those product prices that are actually remembered by many consumers. As we argue in our paper, increased attention to certain product prices may also intensify price competition for those particular products. Put differently, if the share of informed consumers increases while the share of uninformed consumers decreases as a result of increased media coverage of a particular product range, then price competition in that particular product range should also increase.

Our findings can also be related to the literature on rounding and retail pricing, as we observe and discuss pricing preferences of retailers. Providing valuable insights from a behavioral perspective, in a recent study Strulov-Shlain (2022) analyzes firms' pricing behavior, especially 99-ending prices due to the assumed left-sided bias of consumers in a large retail scanner data sample. He finds evidence that firms' pricing behavior is more consistent with rule-of-thumb and heuristics than with optimization, given the demand structure. In another recent study on 99-ending prices, Conlon and Rao (2020) document a rigidity in retail prices caused by retailers' preference for prices ending in 99 and therefore changing prices in whole dollar increments.

Furthermore, this paper provides empirical insights related to the theoretical discussion on the role of competition for pass-through. In addition to demand elasticity, competitive pressure in a market is also related to VAT pass-through. Weyl and Fabinger (2013) find that pass-through under monopoly can be higher or lower than pass-through under perfect competition, depending on the curvature of demand. However, Fuest et al. (2021) find empirical evidence that in product markets with only a few suppliers, the price reduction induced by a temporary VAT rate cut was less pronounced than in product markets with many suppliers. In the following subsection, we discuss the theoretical background of our empirical analysis.

1.2.2 Theoretical background

Pertinent to our research is foremost the theoretical literature on tax pass-through. Which pass-through rate should we expect? For sufficiently inelastic demand, the pass-through rate should be 100 percent. Under perfect competition, the classical result is that the pass-through rate is affected by the ratio of the elasticities of supply and demand. The higher the ratio of the elasticity of supply to the elasticity of demand, the higher the pass-through rate and the higher the tax burden that falls on the inelastic side of the market (see Weyl and Fabinger, 2013).[7] However, also for imperfectly competitive markets, the pass-through rate should be complete when demand is inelastic. While excess pass-through can occur with imperfectly competitive suppliers, it requires further curvature restrictions on demand. As pointed out by Conlon and Rao (2020), excess pass-through may be based on "unrealistic restrictions on demand curves." Also, Anderson et al. (2001) suggest for differentiated price competition a 100 percent pass-through rate when demand is inelastic.[8] As argued above and following Cotropia and Rozema (2018), we consider the demand for menstrual hygiene products inelastic within common price ranges. We hence conclude that the pass-through rate should be 100 percent.[9]

1.3 Chronology, data, and empirical strategy

1.3.1 Chronology

Our data set covers the time period with all stages relevant to the VAT reduction. Starting with week number (WN) 42 (10/14/2019), the data set includes the week when the draft law was released (WN 43, 2019). Also, the week when the final vote in the German parliament, the Bundestag, took place (Thursday of WN 45, 2019) and the week when the second chamber, the Bundesrat, approved the new tax law (Friday of WN 48, 2019) are included. On January 1, 2020 (WN 1, 2020), the legislative amendment officially came into force and the applied VAT rate was 7 percent instead of 19 percent as before. A more detailed chronology can be found in Appendix 1.A.

[7]Pass-through under third degree price distrimination is further analyzed by Miklos-Thal and Shaffer (2021).

[8]See also Genakos and Pagliero (2022).

[9]In retail markets, pass-through rates can also depend on the nature of vertical contracts with manufacturers (see Haucap et al., 2021), this concern does not affect changes in the VAT rate, which are directly applied at the retail level.

1.3.2 Data

The data set contains the gross prices which are also the posted prices for all female hygiene brand products offered by 10 large German retailers,[10] including leading drugstores, discounters, and supermarkets. The retailers included in our data set serve about 60 percent of the demand for hygiene products and 40 percent of the demand for grocery retailing in the German market. The data contains gross prices on a weekly basis for 21 weeks, namely from WN 42 in 2019 and WN 10 in 2020 (inclusive) at the national level.

Feminine hygiene products fall into two main categories: tampons and sanitary pads. Differentiated by the official Global Trade Item Number (GTIN), there are 309 unique products, 114 tampons and 195 sanitary pads. Using the product names, these can be categorized within the product type in terms of the size of the tampons and sanitary pads, as well as the number of items per package and other qualitative product characteristics. More details and descriptive statistics on our data can be found in Appendix 1.B.

1.3.3 Empirical strategy

We conduct an event study defining a dummy-variable named $after$ taking the value 1 for periods after the treatment. We distinguish in our analysis the prices before WN 49 (excluding week 49) in 2019 as the "before" treatment period, and we classify the prices after WN 1 (including week 1) in 2020 as the "after" treatment period.[11] The turn of the year was Tuesday to Wednesday, so most of week 1 was in 2020. Additionally, we define a dummy-variable taking the value 1 for WN 49 to WN 52 in 2019, which are the weeks between the passing of the law and its entering into effect. This variable is named *between* and captures the anticipatory price effect of the VAT change.[12] Altogether, there are seven weekly observations before the identified transition period, four observations transition period, and there are nine weekly observations after the VAT change.

[10]The data were provided by a market participant and therefore anonymized. We thus may not provide further information on their identity, alternative distribution channels, or more narrow characteristics.

[11]As a robustness check we present the baseline and product type result when week 1 belongs to the *between* period in Appendix 1.C.1.2.

[12]Appendix 1.C.1.3 presents a robustness check when week 48 belongs to the *between* period.

We work with the before-and-after distinction to determine the relative price change triggered by the official decision to reduce VAT and its introduction. Our approach can be formalized as follows:

$$price_{i,t} = \beta_0 + \beta_1 \cdot between_t + \beta_2 \cdot after_t + a_i + u_{i,t} \tag{1.1}$$

with $price_{i,t}$ representing the logarithmic prices of different identified retailer-product combinations i fluctuating over time t. The variable $between_t$ indicates the anticipatory effect of the passing of the law and $after_t$ the VAT reduction as a before-and-after dummy. We control for retailer-product fixed effects a_i.[13] The two dummy variables are interacted with others in the further analysis.

1.4 Results

We begin with a general overview of the effect of the VAT change and contrast it to the general CPI. We then provide a detailed analysis of the differential effects of product type, store type, and competition (given by the number of retailers offering the products). Further analyses can be found in Appendix 1.C.3, 1.C.4 and 1.C.5. Unless otherwise stated, we perform panel regressions that include retailer-product fixed effects and report bootstrapped standard errors in our analysis. Furthermore, all tests reported are post-regression Wald tests with the corresponding p-value unless otherwise indicated.

1.4.1 Main results

Figure 1.1 shows the prices of the menstrual hygiene products over time, indicated changes are relative to $t = -11$. It shows the important interweekly transitions. The gross price effect as well as the anticipatory effect of the VAT reduction are evident. The figure shows the immediate reaction of the retailers to the approval of the law from $t = -5$ to $t = -4$.[14] Even before the official date of the VAT reduction, prices started to drop. As expected[15], the policy change was finalized on the Friday of WN

[13] Table A.7 in 1.C.1.5 compares our baseline regression for different fixed effects.

[14] See also Klug in Lebensmittelzeitung, an industry portal for the grocery retail and consumer goods industry focusing on Germany, of Dec 10, 2019.

[15] The agenda for the meetings of the Bundesrat is announced 10 days in advance.

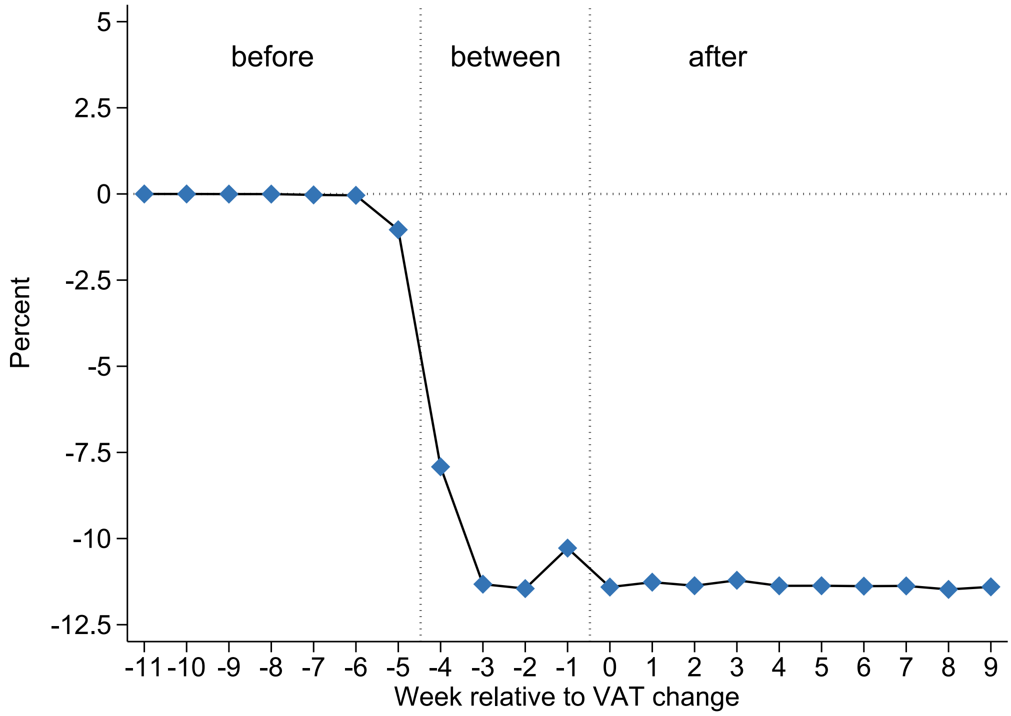


Figure 1.1: This graph shows the price changes of menstrual hygiene products over time. Price changes are relative to $t = -11$. $t = 0$ represents the first week of 2020 when the law came into effect. $t = -5$ marks week number 48 in 2019 when the law was passed.

48 which equals $t = -5$ in Figure 1.1. The VAT reduction was effective by $t = 0$.

This analysis is based on the time series of prices for treated products, with no untreated product group as a control group. The implicit counterfactual is the pre-reform price of the treated product group. Using this counterfactual seems reasonable because prices were very stable in the pre-reform period (i.e., the period before the law was passed).

A price reduction of 10.084 percent corresponds to a 100 percent pass-through. Since the prices in the data set include VAT, the tax reduction for menstrual hygiene products from 19 to seven percent corresponds to a gross price reduction of 10.084 percent for each product affected by the VAT cut.[16] Thus, the benchmark for the full pass-through is 10.084 percent.

When we adopt a log-level approach, we can interpret the estimated coefficients

[16]The gross price P is $P_k = (1 + s_k) \cdot p$, $k = 0, 1$, with p reflecting the net price of the product and s_k being the VAT rate. Before the VAT reduction the tax rate is $s_0 = 0.19$ and after the law came into effect the new tax rate $s_1 = 0.07$ applies. The new gross price equals 89.916 percent since $P_1/P_0 = 0.89916$. The former gross price has been reduced by 10.084 percent.

Table 1.1: Baseline regression.

	All products
	ln(price)
between	-0.114***
	(0.000573)
after	-0.119***
	(0.000318)
Constant	1.031***
	(0.00274)
$H_0 : after = -0.1063$	$p < 0.001$
Retailer-Product FE	✓
R^2	.9978904
Observations	17015

Notes: Bootstrapped standard errors are reported in parentheses. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. The H_0 row shows the Wald test for full pass-through. The null hypothesis is that the effect is equal to the benchmark for full pass-through. The benchmark is -0.1063. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(β_j) for dummy variables only as an approximation of the true marginal effect. The formula $(e^{\beta_j} - 1) \cdot 100$ gives the exact percentage change. Therefore, an estimated coefficient of -0.1063 corresponds to our benchmark of a gross price reduction of 10.084 percent. Estimated coefficients smaller than -0.1063 show an excess pass-through while estimated coefficients larger than -0.1063 reflect a less than full pass-through.

Table 1.1 shows the baseline regression as described in the Empirical strategy subsection 1.3.3. Both the anticipation effect captured by the variable *between* and the tax incidence captured by the variable *after* are highly significant for the full sample. Both are highly significant above the benchmark of full pass-through. The estimated price reduction after the VAT change is 11.23 percent[17] which indicates a 1.15 percentage-point excess reduction which can be translated to a 11.4 percent excess pass-through.[18] The price reduction increased significantly *after* the VAT change compared to the transition period *between*.

Remarkably, prices are already reduced in the *between* phase by more than the

[17]Note that $(e^{-0.119} - 1) \cdot 100 = -11.23$.
[18]This is calculated as follows: $1.15/10.084 = 0.114$.

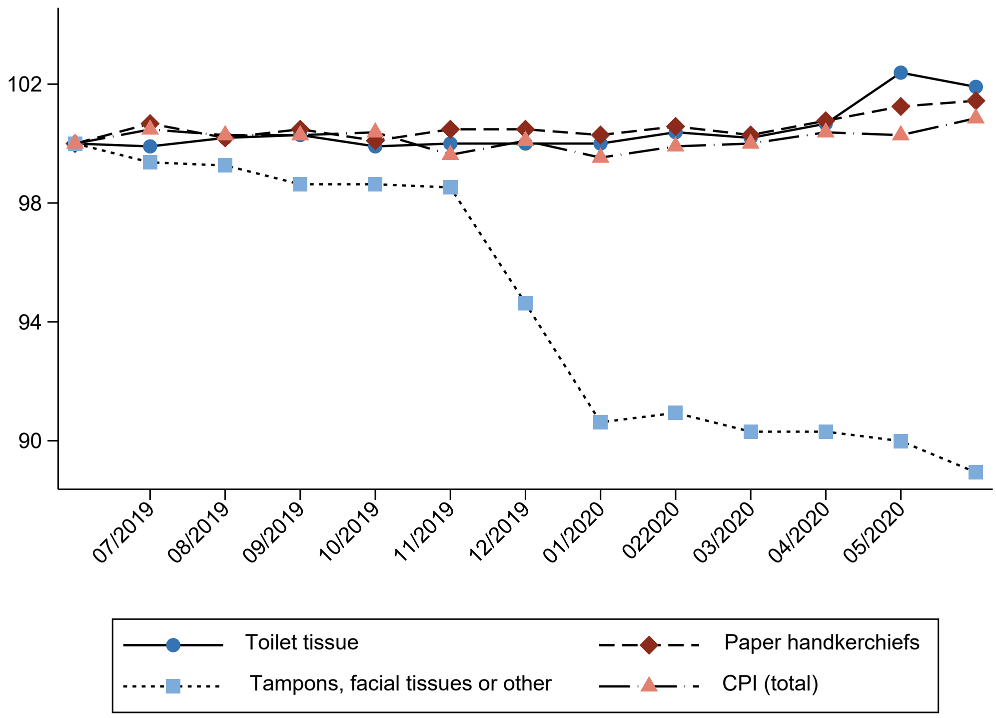


Figure 1.2: This figure shows the Consumer Price Index (CPI) for assorted consumer goods for personal care. The triangle line shows the overall CPI. The line with square indicator represents the index including menstrual hygiene products. The dot and diamond line represent categories that rely on similar input but were not affected by the VAT change. The law to change VAT was passed in late November 2019. The VAT change came into effect in January 2020. Base 06/2019; Data source: (Statistisches Bundesamt (Destatis), 2022).

expected VAT change suggests. The estimated price reduction is 10.77 percent and reflects an excess price reduction of 0.69 percentage points.

The observed price drop is not the result of a general price shock, as Figure 1.2 shows. While similar products with respect to primarily used inputs remain at the same price level as before or even show upward tendencies, the line with a square indicator in Figure 1.2 shows a clear drop between November 2019 and January 2020.[19] We conclude that the price decrease is almost certainly the result of the VAT reduction for this product group.

[19]The selected categories are: Tampons, facial tissues, and other sanitary products which include menstrual hygiene products, and then toilet tissues and paper handkerchiefs. The products of these categories are mainly cellulose-based. Therefore, we can exclude an overlaying price trend caused by input factor prices for the tampon, facial tissues, and other sanitary products category. Further, a large share of the products in these categories can be assumed to be necessities. Besides the similarity in input factors, demand for these products is rather stable.

Table 1.2: Difference-in-differences estimation results.

Variables	value
Diff-in-diff	-9.548**
	(0.412)
Constant	100.2**
	(0.168)
Observations	36
R^2	0.979
Mean control t(0)	100.2
Mean treated t(0)	99.07
Diff t(0)	-1.115
Mean control t(1)	100.8
Mean treated t(1)	90.18
Diff t(1)	-10.66

Notes: Control group: Toilet tissue and paper handkerchiefs. Treated: Tampon, facial tissues or other. $t(0)$: 06/2019 - 11/2019; $t(1)$: 01/2020 - 06/2020. 12/2019 is excluded (anticipatory efffect after passing of the law in November 2019). Means and standard errors are estimated by linear regression. Standard errors are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1.2 shows the results of a difference-in-differences estimation of the CPI data to size the effect. The treatment effect is significant (p-value<0.01) and the estimated size of the reduction is not significantly different from the size of the VAT reduction (p-value=0.202).

1.4.2 Heterogeneity

This section analyzes the heterogeneous price effects within the data. We distinguish between the two product types, between the two different store types, and between the degree of competition (one retailer vs. more than one retailer offering the product). Table 1.3 presents the regression results and post-regression tests evaluating the estimates against the full pass-through benchmark.

Table 1.3: Heterogeneities.

	Product type	Store type	Competition
	ln(price)	ln(price)	ln(price)
between	-0.124***	-0.108***	-0.101***
	(0.000648)	(0.000565)	(0.00195)
after	-0.127***	-0.111***	-0.106***
	(0.000470)	(0.000369)	(0.00102)
between × tampon	0.0250***		
	(0.00115)		
after × tampon	0.0181***		
	(0.000542)		
between × drugstore		-0.00967***	
		(0.00108)	
after × drugstore		-0.0143***	
		(0.000611)	
between × competition			-0.0143***
			(0.00204)
after × competition			-0.0147***
			(0.00106)
Constant	1.031***	1.031***	1.031***
	(0.00274)	(0.00274)	(0.00274)
$H_0 : after = -0.1063$	$p < 0.001$	$p < 0.001$	$p = 0.49$
$H_0 : after + after \times interaction = -0.1063$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Retailer-Product FE	✓	✓	✓
R^2	.9980533	.9979643	.9979146
Observations	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. In the Product type column, *between* and *after* show the price reduction for sanitary pads; × *tampon* shows the difference for tampons. In the Store type column, *between* and *after* show the price reduction for other stores; × *drugstore* shows the difference for drugstores. In the Competition column, *between* and *after* show the price reduction without competition; × *competition* shows the difference with competition. The H_0 rows show the Wald tests for full pass-through. The null hypotheses are that the effect is equal to the benchmark for full pass-through. The benchmark is -0.1063.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.4.2.1 Product-type differences

With sanitary pads and tampons, we observe two different product types in our data. We have around 58 percent sanitary pads and 42 percent tampons. In Table 1.3 the first column shows the regression results for the two different product types. For sanitary pads we find a significantly larger price reduction than a full pass-through *after* the VAT change. For tampons the price reduction *between* the passing of the law and the coming into effect was more restrained and around one percentage point below 10.084 percent, but the estimated effect *after* the VAT change exceeded the full pass-through benchmark significantly. We find a significant difference between the two product types.

1.4.2.2 Store-type differences

Menstrual hygiene products belong to the drugstore category. Although also supermarkets offer products from this category, only drugstores offer a greater variety within the individual product groups. We observe that around 55 percent of the data are from drugstore-type outlets.

Table 1.3 the second column shows the regression results for drugstores compared to other stores in column Store type. The price reduction of other stores exceeds the full pass-through benchmark of 10.084 percent significantly *after* the VAT change. Drugstores reduce their prices by around one percentage point more in the *between* phase and by around 0.5 additional percentage points when compared to other stores *after* the VAT change. Both additional effects are significant.

1.4.2.3 Retailer competition

We measure competition with respect to the number of retailers offering a certain product. Products offered by just one retailer are referred to as without competition while products which we find in two or more retailers are called with competition. In our data, we classify around 2.5 percent as products without competition. Products without competition are presumably niche products, as fewer retailers are interested in selling them; they must not be confused with products sold by monopolies.

Table 1.3 the third column shows the regression results when we distinguish between products with and without competition. The estimated price reduction of a product without competition *after* the VAT change is not significantly different from the 10.084 percent benchmark. The price reduction related to competition is highly significant *after* the VAT change and the reduction for products with competition significantly exceeds the benchmark of 10.084 percent by around 1.4 percentage points.

Since the data set does not cover the entire supply-side, this number must be interpreted as a proxy for the true number of retailers. In fact, some of the products stocked by only one retailer in our sample might actually be offered by more than one retailer in the whole market. As a result, the estimate when only one retailer offers a product can be interpreted as an upper bound of the relative price effect. Since we observe a smaller price change for our one-retailer estimate, the real difference between one retailer and more than one retailer stocking a product may be larger than our results suggest.

1.5 Explanations for excess pass-through

1.5.1 Net price reduction

The excess pass-through identified in the results section implies a reduction in net prices, revealing the true costs borne by retailers. Figure 1.3 shows the changes in net prices over the observed period. In the interim period[20], net prices were reduced by about 11 percentage points, which translates directly into opportunity costs for retailers. After the VAT change net prices remain around one percentage point below the earlier price level.

In the following section, we discuss the main channel of the excess pass-through rate: rounding effects. In a subsequent section, we explain what drives this effect and how it interacts with the identified heterogeneities.

1.5.2 Rounding

A substantial literature indicates that retailers have a general preference for (gross) prices to end with a 0, 5 or 9 in the second decimal.[21] The retailers in our data are no exception: we observe these second decimals for more than 98 percent of all price observations both before and after the tax reduction. Six retailers set all prices ending with one preferred second decimal while four retailers use two or all three

[20]We refer to this as the *between* phase in our regressions.

[21]These preferences result from the so-called "pricing in the nines" phenomenon (see Basu, 1997) which uses the idea that consumers ignore the last digits of a price (see, e.g., Nagle and Holden, 1987) and has been observed for decades (see, e.g., Ginzberg, 1936, who reported this phenomenom as "customary pricing"). Effects of this pricing strategy have been studied e.g., in Schindler and Kibarian (1996) Schindler and Kirby (1997), and Stiving and Winer (1997). Basu (1997) provides an economic explanation for this pricing strategy. Strulov-Shlain (2022) added just recently to this literature, providing empirical evidence against this strategy from a retailer's point of view, as lost profits by setting .99 instead of the closest integer price exceed the revenue from additional demand.

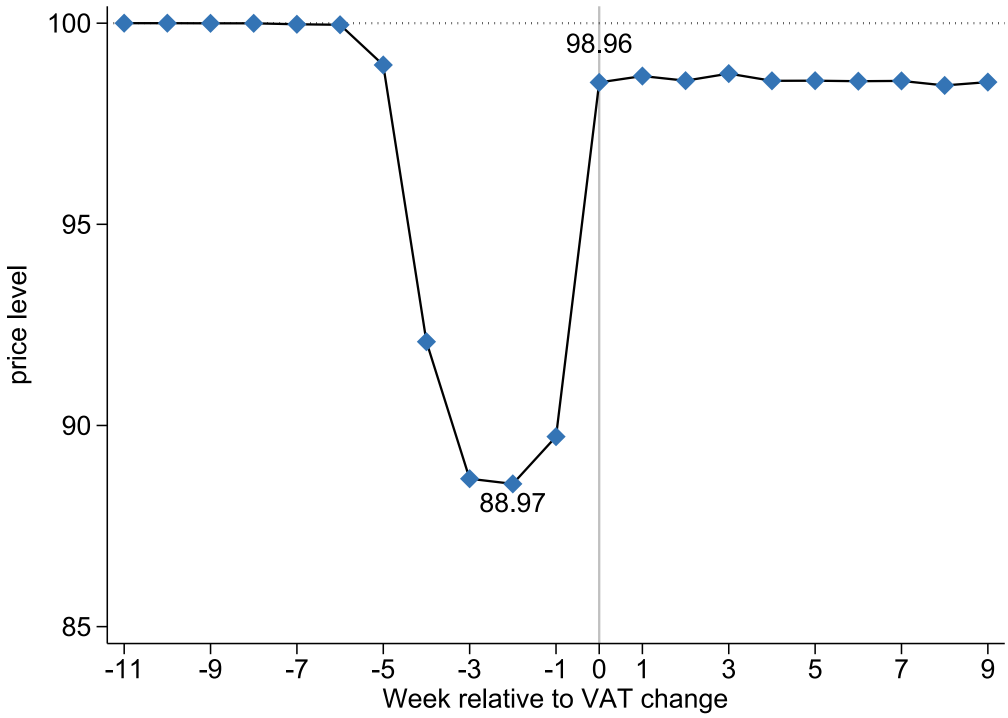


Figure 1.3: This figure shows the normalized net price (without VAT) evolution over time. $t = -5$ marks week number 48 in 2019 when the law was passed. $t = 0$ represents the first week of 2020 when the law came into effect. The price level is normalized to $t = -11$.

common decimals.

Preference for specific second decimals demands adjustments of net prices to match the targeted gross price when the VAT changes, and we observe almost exclusively downward-adjusted net prices, namely in 91.55 percent of retailer-product combinations.[22] Typically, retailers' net price adjustments have leeway in either direction: Compared to the benchmark when the new VAT is mechanically applied, they can round down or up. What they almost always do in our data is rounding down.

To further illustrate the rounding-down effect, Figure 1.4 shows the distribution of differences between the expected and actual gross price after the VAT change for the observed retailer-product combinations.[23] The mean difference of these prices is 2.93 euro cents. The figure also shows a wide range of realized price differences.

[22]We calculated expected prices after the VAT change based on the median price before the law was passed. Then we calculated the difference between this expected price and the actual price after the VAT change.

[23]The expected gross price results when the new VAT is applied without changing the net price.

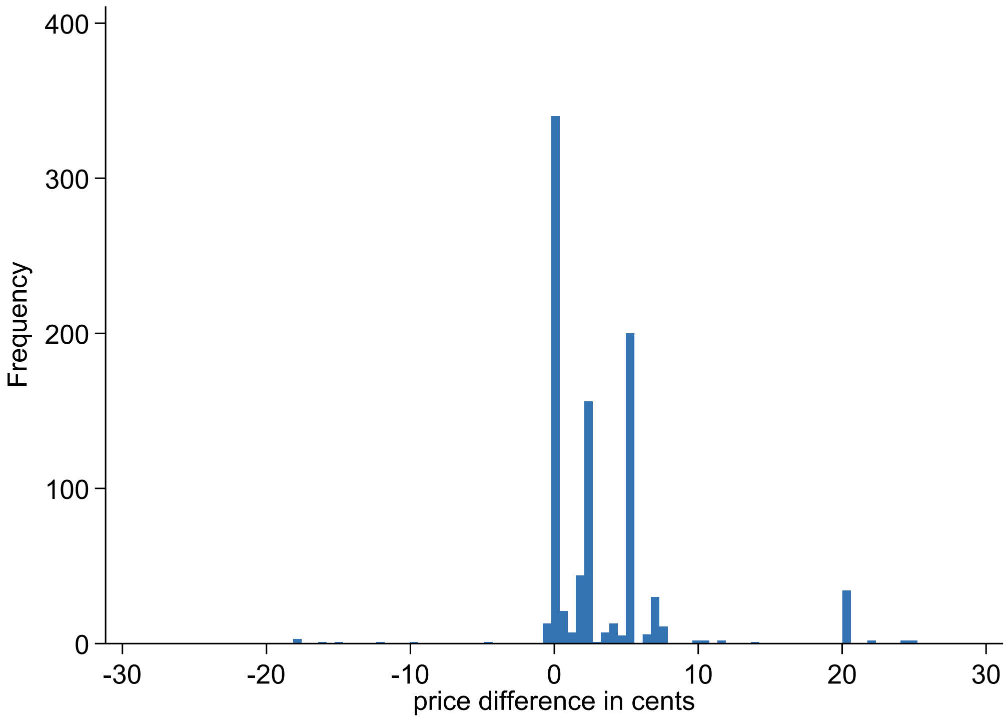


Figure 1.4: This figure shows the differences between the expected price after the VAT change (based on the median price before the law was passed) and the actual median price after the VAT change in euro cents for retailer-product combinations. A negative sign indicates an upward effect which means the actual price was higher than the expected while a positive sign indicates excess pass-through. Two outliers were removed (-86.07 and -40.24) to improve readibility.

Values below zero indicate a less than full pass-through of the VAT change. The corresponding net prices have been adjusted upwards. Price differences between 0 and 10 euro cents can be explained by rounding to a specific second decimal place. Most of the values in the Figure 1.4, 86.72 percent, fall within this range.

Crucially, retailers round not just to one of the usual second decimal places, but to the same decimal place as before the VAT change, which we call consistent rounding. Indeed, this is the case for 86.28 percent of the observations. We focus on this most common strategy in the following.

Given the rounding to the same second decimal regularity, the maximum difference between the actual and rounded gross price is 9.99 euro cents. As an example, assume a gross price of 4.99 euro including a VAT of 19 percent.[24] The corresponding net price is 4.99 euro/1.19 $\approx$ 4.1933 euro. With the new VAT rate of 7 percent, the new

[24] In our data, prices range from 0.75 euro to 6.45 euro, as shown in the descriptives table in the Appendix 1.B.

gross price would be 4.4868 euro. Now, if this gross price were rounded down to a 9 in the second decimal, the new gross price would be 4.39 euro, which means a gross price difference of 9.36 euro cents. Obviously, the corresponding net price also changes to 4.1028 euro, which is an absolute reduction of the net price by 9.05 euro cents. The gross price in this example has been reduced from 4.99 euro to 4.39 euro, a reduction of 60 euro cents. This is a relative reduction of 12.03 percent, which is 1.94 percentage points higher than the 10.084 percent. This difference in relative price reduction varies for different prices and may increase for smaller gross prices, as the maximum difference between the actual and rounded gross price remains at 9.99 euro cents.

We now estimate the effect of the "rounding to the preferred decimals" strategies and their magnitudes in our data. Table 1.4 shows the results, and the "rounding only" dummy equals one only if a retailer uses the strategy of rounding the price of a product to one of the most preferred decimals, namely 0, 5, and 9, before and after the VAT change without consistent rounding, while the "consistent rounding" dummy equals one if a retailer uses the same of the most preferred decimals before and after the VAT change.[25]

First, we find that for retailer-product combinations without these rounding strategies (no rounding), the price reduction of 9.95 percent *after* the VAT change is not significantly different from the benchmark (p-value=0.805). For these, we find full pass-through, but no excess pass-through. For prices that are rounded, but not consistently (rounding only), we find no significant difference from prices that are not rounded (p-value=0.308). When prices are consistently rounded to the same decimal as before the VAT change (consistent rounding), they are significantly reduced by 1.55 percentage points more than those without rounding which adds up to a reduction of 11.49 percent and is a highly significant excess reduction. Thus, the rounding strategy significantly amplifies the price reduction.

We conclude that the observed excess pass-through is mainly driven by rounding. In particular, the strategy of rounding to the same second decimal (consistent rounding) contributed to excess price reduction.

[25] The proportions of these three groups are as follows: Consistent rounding: 86.28 percent, rounding only: 13.28 percent, no rounding: 0.44 percent.

Table 1.4: Rounding and consistent rounding effects.

	Rounding
	ln(price)
between	-0.117***
	(0.00516)
after	-0.105***
	(0.00627)
between × rounding only	0.00776
	(0.00524)
after × rounding only	-0.00643
	(0.00631)
between × consistent rounding	0.00275
	(0.00518)
after × consistent rounding	-0.0156*
	(0.00626)
Constant	1.031***
	(0.00274)
$H_0 : after \; = \; -0.1063$	$p = 0.8053$
$H_0 : after \; + \; after \times rounding\ only \; = \; -0.1063$	$p < 0.001$
$H_0 : after \; + \; after \times consistent\ rounding \; = \; -0.1063$	$p < 0.001$
Retailer-Product FE	✓
R^2	.9979059
Observations	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the rounding effect when prices are rounded to 0, 5, or 9 in the second decimal (denoted by × *rounding only*) and the rounding effect when prices are consistently rounded to the same second decimal (denoted by × *consistent rounding*). There is no overlap between these two segmentations. *between* captures the price reduction due to the anticipation effect; *after* shows the price reduction after the VAT change. The H_0 rows show the Wald tests for full pass-through. The null hypotheses are that the effect is equal to the benchmark for full pass-through. The benchmark is -0.1063. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.5.3 Media coverage

The question remains as to why retailers largely round down, and why they do the opposite in some cases. We argue that media attention led to this behavior.[26] We provide both anecdotal evidence and quantitative data on this issue.

As for anecdotal evidence, there are press articles and retailers' promotional efforts. These promotion activities are mainly observed in the transition phase and were linked to the large net price reduction in this phase. News articles covering the VAT change were published not only in tabloids, but also different international, national, and regional newspapers.[27] The variety of outlets suggests a broad coverage, reaching different groups of consumers. Around the same time, retailers were actively promoting the VAT cut and the corresponding price reduction even before the tax change was observed.[28] This is a substantial amount of attention for a product that under other circumstances would hardly even be noticed by the media.

To further quantify the media attention, we systematically searched for press articles and Twitter posts on the topic. The data sets result from own research using free online sources like Google News and scraping Twitter data. Figure 1.5 shows data on various (online) press articles covering the topic of VAT reduction for menstrual hygiene products as well as tweets posted on Twitter dealing with the topic.

There are two points in time when both press articles and tweets divert from their previous trend. The first point in time is connected to the announced intention of the Ministry of Finance to reduce the applied tax rate for menstrual hygiene products by January 1, 2020. Both tweets and also, somewhat later, press articles resulted in an increase of attention reflected by a higher number of tweets and articles. Thereafter,

[26] Hindriks and Serse (2022) also argue that their analyzed tax cut was widely covered in the media and therefore firms were publicly pressured to pass through the entire VAT cut. Relatedly, the size differences of pass-through rates between Montag et al. (2021) and Dovern et al. (2022) are presumably attributable to the greater or more direct focus on fuel price tax cuts in the latter study.

[27] See, for example, online sources from: Augsburger Allgemeine (Nov 8, 2019), Gala (Nov 8, 2019), The New York Times (Nov 12, 2019), The Washington Post (Nov 12, 2019), tagesschau (Nov 29, 2019), RTL (Dec 4, 2019), Mitteldeutsche Zeitung (Dec 4, 2019), Handelsblatt (Dec 12, 2019), all retrieved on March 15, 2023.

[28] See, for example, press releases from dm (Dec 2, 2019), Lidl (Dec 5, 2019), EDEKA (Dec 9, 2019), Netto (Dec 9, 2019), all retrieved on March 15, 2023. Similarly, promotion activities were observed for other VAT cuts, see e.g., Crossley et al. (2009) who provide anecdotal evidence on active retailer promotion on a temporary VAT cut in the UK.

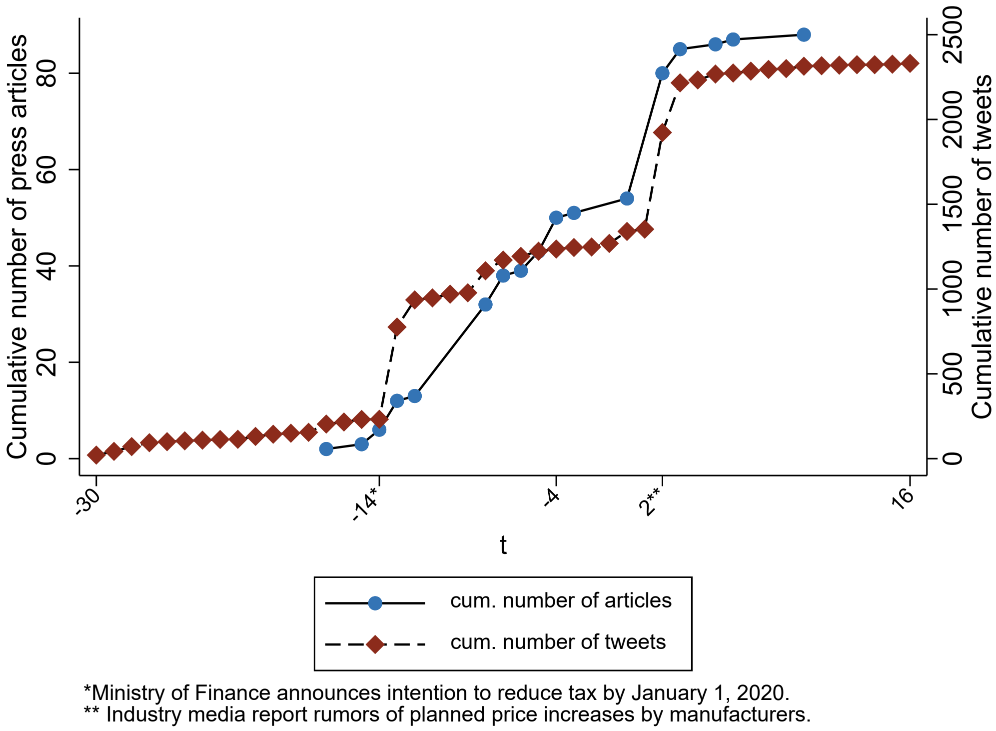


Figure 1.5: This figure shows the evolution of media coverage on the VAT reduction for menstrual hygiene products. The diamonds represent the cumulative number of tweets including either the words “tampontax” and “Germany” combined or the German word “tamponsteuer” starting WN 23 in 2019 until WN 17 in 2020. The dots represent the number of (online) press articles covering the topic from summer 2019 until late spring 2020. These time horizons frame the period of our price observations generously.

the number of tweets remained on a slightly increased trend with only one small level-shift when an ongoing progress in the legislative process could be observed. Also, press articles covered this period on a moderate level.

The second point in time is around two weeks after the VAT reduction came into effect. In WN 3 of 2020 an industry portal for grocery retail and consumer goods industry focusing on Germany reported on rumors of planned price increases by manufacturers.[29] These rumors lead to a media response of similar size as the announcement of the Ministry of Finance did before. Both press articles and tweets increased again on this negative news from a consumer perspective. This supports the assumption about media coverage and attention and can be interpreted as an indication that retailers would also have been affected by negative publicity if they

[29]The factory selling price between manufacturer and retailer is usually negotiated in annual meetings (see e.g., Kaas, 1993; Bundeskartellamt, 2014).

had not passed through the entire reduction.

The intense media coverage of this VAT reduction may have influenced consumer attention and thus shifted the price competition between retailers towards tampons and sanitary pads, as discussed above. If media coverage draws consumers' attention to these products, supermarkets and drugstores can be expected to compete more vigorously in these product lines to attract customers to their stores. On the other hand, lowering prices on products to which consumers pay little or no attention is not attractive to retailers, because retailers lose mark-ups on these products without attracting enough additional customers to their stores. Therefore, the best solution is to lower prices and forgo margins on products that consumers pay significant attention to. In addition, any media report of a retailer passing through less than 100 percent could have resulted in negative publicity, especially given the highly political campaign surrounding the tax cut, which suggested that the previous VAT rate negatively discriminated against women. Therefore, retailers rounded down prices to avoid negative press coverage and implemented a larger price reduction than the VAT change would have implied.

To sum up, the attention in the form of press reports, tweets, and advertisements created pressure and expectations for the pass-through of the VAT cut, especially through the claim of negative discrimination used to support the campaign. While we cannot provide evidence of the counterfactual — that passing through less than 100 percent would have led to a loss of sales — we firmly conjecture that media coverage and attention did lead to the high share of rounding-downwards discussed in section 1.5.2.

1.5.4 Excess pass-through in "competitive segments"

In the previous sections, we established that excess pass-through is primarily a result of the rounding-downward effects triggered by media coverage and attention. It remains for us to demonstrate how the use of the rounding strategy is related to the heterogeneities in pass-through rates shown in the results section.

Table 1.5: Rounding and heterogeneities.

(1) Product type × Rounding			(2) Store type × Rounding			(3) Competition × Rounding		
	Freq.	Percent		Freq.	Percent		Freq.	Percent
Sanitary pads			*No drugstore*			*No competition*		
Rounding=0	4	0.76	Rounding=0	4	0.97	Rounding=0	1	1.41
Rounding=1	521	99.24	Rounding=1	409	99.03	Rounding=1	70	98.59
Total	525	100	Total	413	100	Total	71	100
Tampons			*Drugstore*			*Competition*		
Rounding=0	0	0	Rounding=0	0	0	Rounding=0	3	0.36
Rounding=1	386	100	Rounding=1	498	100	Rounding=1	837	99.64
Total	386	100	Total	498	54.67	Total	840	100
Total			Total			Total		
Sanitary pads	525	57.63	*No drugstore*	413	45.33	*No competition*	71	7.794
Tampons	386	42.37	*Drugstore*	498	54.67	*Competition*	840	92.21
Total	911	100	Total	911	100	Total	911	100

Notes: This table shows descriptive insights into the heterogeneities of retailer-product combinations in the 'Total' row and descriptive insights within the competitive segments conditioned on whether retailers round to the most common second decimals ($Rounding = 1$).

Tables 1.5 and 1.6 show the details. Table 1.5 provides descriptive insights into these heterogeneities when we condition on whether retailers round to the most preferred second decimals or not. The regression results presented in Table 1.6 show the impact of store type, product type, and retailer competition when we control for rounding.

Column (1) of Table 1.6 includes the product type heterogeneity while Table 1.5 column (1) shows the number of observations for the two product types differentiated by rounding. We observe that close to all sanitary pad products and all tampon products round. The regression results show a highly significant higher price reduction for sanitary pads than for tampons when both round during the $after$ phase of the VAT change.[30] This finding is consistent with the results from section 1.4.2.1, which also show a larger reduction for sanitary pads. We conjecture that sanitary pads are the more "competitive" product as there are more variants.

The second column of Table 1.6 presents the store-type heterogeneity. Table 1.5 column (2) shows that all drugstore products are rounded. The regression results indicate that drugstores reduce their prices more than other grocery stores by a highly

[30]The estimated difference is about 1.8 percentage points after the VAT change.

significant margin, given that both round *after* the VAT change.[31] This finding is consistent with section 1.4.2.2 and is likely due to menstrual hygiene products being associated more with drugstores, which makes them a competitive core product for these retailers, but less so for regular supermarkets.

The competition heterogeneity is included in column (3) of Table 1.6. Table 1.5 column (3) shows the number of observations with and without competition differentiated by rounding. In the regression results, we see highly significantly more price reductions with competition both before and after the VAT change.[32] This result is consistent with the result in section 1.4.2.3 and is also intuitively consistent. We divide the products into two groups: with and without competition. The term "without competition" should not be confused with a monopolistically supplied product. In our context, "without competition" means that we are talking about special niche products that are offered by only one retailer and therefore receive less media attention, while the products with competition are arguably more popular. Thus, it is not surprising that, once again, the pressure of media attention had a greater effect on the products with competition than on those without.

We conclude that the excess pass-through we observe in certain subsamples is fully consistent with the rounding effect. We still observe differences across competitive segments. In Appendix 1.C.2 we repeat this analysis when we condition on consistent rounding and excess pass-through.

[31] The estimated difference is around 1.4 percentage points after the VAT change.

[32] The estimated difference is about 1.4 percentage points after the VAT change.

Table 1.6: Products rounding interacted with

	+ Product type	+Store type	+Competition
	ln(price)	ln(price)	ln(price)
between	-0.117***	-0.117***	-0.110
	(0.00526)	(0.00526)	(0.0000)
after	-0.105***	-0.105***	-0.0832***
	(0.00575)	(0.00575)	(0.0106)
between × rounding	-0.00709	0.00907	0.00962***
	(0.00526)	(0.00522)	(0.00177)
after × rounding	-0.0222***	-0.00613	-0.0227*
	(0.00574)	(0.00569)	(0.0106)
between × tampon	0.0250***		
	(0.00123)		
after × tampon	0.0182***		
	(0.000605)		
between × drugstore		-0.00977***	
		(0.00115)	
after × drugstore		-0.0143***	
		(0.000628)	
between × competition			-0.0102
			(0.00719)
after × competition			-0.0287*
			(0.0127)
between × rounding × competition			-0.00414
			(0.00738)
after × rounding × competition			0.0143
			(0.0128)
Constant	1.031***	1.031***	1.031***
	(0.00271)	(0.00273)	(0.00268)
$H_0 : after = -0.1063$	$p = 0.7882$	$p = 0.7882$	$p < 0.05$
$H_0 : after + after \times rounding = -0.1063$	$p < 0.001$	$p < 0.001$	$p = 0.6524$
$H_0 : after + after \times interaction = -0.1063$	$p < 0.001$	$p < 0.05$	$p = 0.4228$
Retailer-Product FE	✓	✓	✓
R^2	.9980563	.9979651	.9979165
Observations	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the different heterogeneities interacted with a dummy identifying prices with rounding to the most common second decimals (denoted by × *rounding*). In column +Product type *between* and *after* show the price reduction for sanitary pads; × *tampon* presents the difference for tampons. In column +Store type *between* and *after* show the price reduction for other stores; × *drugstore* presents the difference for drugstores. In column +Competition *between* and *after* show the price reduction without competition; × *competition* presents the difference with competition. The H_0 rows show the Wald tests for full pass-through. The null hypotheses are that the effect is equal to the benchmark for full pass-through. The benchmark is -0.1063. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.6 Conclusion

In this paper we studied the price effects caused by a VAT (value-added tax) reduction for menstrual hygiene products in Germany. The policy change was initiated by a petition in the parliament and campaigns, highlighting that these products are subject to the regular (highest) VAT rate, but are purchased only by menstruating women, thus suggesting a negative discrimination. Exploiting a rich set of retail prices, we analyze the price level before the parliament's decision and after the implementation of the VAT reduction (plus a between phase). As demand for these products should be rather inelastic within certain ranges, we expected a 100 percent pass-through.

Our analysis shows not only a complete pass-through of the VAT reduction, but even excess pass-through. In our baseline regression, we found a 1.15 percentage-point excess price reduction which corresponds to a 11.4 percent excess pass-through. We also find the anticipation effect to be already similar in magnitude, as retailers cut prices as soon as the law is passed rather than waiting until it goes into effect. In further regressions decomposing for different "competitive segments" of the market, we found substantial differences between subsamples: A higher excess pass-through in sanitary pads compared to tampons, in drugstores compared to grocery stores, and for more popular products which are stocked at more than one different retailer. The reduction, however, was at least fully passed through in all segments.

We cannot fully claim causality in the sense that without the passage of the law, the price level would have remained the same as before the reform. However, we are not concerned about a lack of rigor here, because the price change is sharp, coincides with the (passage of the) reform, and the price level is relatively stable before and after. And there is no such price drop for similar products not affected by the reform.

As for the channel of the excess pass-through, we find that the major part of it can be explained by "rounding to preferred decimals" effects. We have identified that almost all prices end with a 0, 5 or 9. In addition, a large proportion of prices stick to the same second decimals after the VAT reduction, which we identified as the most common rounding strategy in our data. The aforementioned differences in competitive segments persist when we control for rounding to the common decimals.

We relate the downward direction of rounding to the preferred second decimals pricing to evidence of media coverage and increasing attention to the VAT reduction. Thus, our paper adds to the recent literature on firm pricing and rounding to specific (second) decimals.

Our findings support the argument that VAT policies intended to benefit the consumers can be successful. In policy debates about such measures, there are frequent criticisms that firms do not pass through tax cuts or negatively factor subsidies to consumers into their pricing. We find no evidence of this. Our findings also suggest that media can play an important role in shifting attention toward the campaign and the products.

Appendices

Appendices

1.A Detailed chronology

The timeline in Table A.1 highlights the important stages of the decision process toward the VAT reduction. Although the process that started with the online petition began well before the first draft, the relevant stages are within our observed time period. The data set used in our analysis starts with WN 42 (10/14/2019) and therefore includes the week when the draft law was released. Also, the moment of the final vote in the German Bundestag in WN 45 (11/07/2019), the date when the legal process was approved in the Bundesrat on 11/29/2019 (Friday of WN 48 in 2019), and WN 1 in 2020 (01/01/2020), the date when the legislative amendment officially came into force, are within the observed time frame.

Table A.1: Timeline from the petition to the passed law.

In 2018	Start of the petition “Die Periode ist kein Luxus” (“The period is not a luxury”) on the platform change.org. (Non-binding for the parliament)
9^{th} *February* 2019	Date of submission of a petition to reduce the VAT for menstrual products in the Bundestag.
8^{th} *March* 2019	The tax reduction topic is in the media and a politician from the conservative ruling party speaks positively about a tax reduction demand.
16^{th} *April* 2019	The guerilla marketing activity named “The Tampon Book” was launched and the campaign started attracting more and more politicians to its side.
27^{th} *May* 2019	The petition in the Bundestag reaches the quorum of at least 50,000 signatures.
3^{rd} *September* 2019	A federal state requests that the Bundesrat deal with the issue at the earliest possible date.
4^{th} *October* 2019	The Minister of Finance announces the intention to reduce the tax as of January 1, 2020.
22^{nd} *October* 2019	The draft law on the tax reduction of menstrual products is fixed.
7^{th} *November* 2019	The Bundestag approves the draft law.
29^{th} *November* 2019	The Bundesrat approves the change in the tax law as proposed by the Bundestag before.
1^{st} *January* 2020	Menstrual hygiene products are taxed 7 percent instead of 19 percent VAT.

1.B Supplementary descriptives

We received our data set from a market participant and have anonymized the included retailers. The 10 retailers can be categorized as either drugstores or mainly grocery stores (referred to as “other stores”). These retailers account for over 60 percent of the annual sales in the German food retail sector.[33]

Our data set includes brand products from two product groups – tampons and sanitary pads – that are classified by a Global Trade Item Number (GTIN). These products come in various sizes and pack sizes, offering a wide range of options. Among the 114 tampon products, 97 (85.1 percent) are from the leading manufacturer brand (LMB), while the remaining 17 are from six other brands. Similarly, out of the 195 sanitary pad products, 168 (86.2 percent) are from the leading manufacturer brand, with the remaining 27 from six other brands. The share of retailers offering only one brand, the LMB, in stock is 40 percent in our data.

In Table A.2, we provide descriptive statistics on the prices of these products based on the competitive segments that we identified. We differentiate between product types, store types, and products offered at one or more than one retailer (we refer to this as “without competition” and “with competition”).

We observe that, on average, sanitary pads are priced 53 euro cents lower than tampons. Both product groups show significant price variation due to the pack size and other characteristics.

We also find that menstrual hygiene products are priced, on average, 29 euro cents lower in drugstores than in other grocery stores.

When we compare prices between products available at multiple retailers (popular products) and those available at only one retailer (more niche products), we find that the average price difference is 93 euro cents.

[33] Our calculation using data from Lebelsmittel Praxis, 2020, which is based on the NielsenIQ/Tradedimensions Top 30.

Table A.2: Descriptive statistics.

	Mean	Median	Std. Dev.	Min	Max	Obs
Total	2.77	3.05	0.96	0.75	6.45	17,015
sanitary pads	2.55	2.89	0.93	0.75	4.79	9,895
tampons	3.08	3.55	0.90	1.19	6.45	7,120
drugstore	2.65	2.89	0.92	0.75	6.45	9,727
other stores	2.94	3.10	0.97	0.75	4.79	7,288
competition	2.70	3.05	0.90	0.75	6.45	15,739
no competition	3.63	4.29	1.16	1.19	6.45	1,276

Notes: This table shows descriptive insights into our data, including the differentiation by competitive segments.

Figure A.1 presents the number of distinct price observations for each product group. This illustrates that more than 60 percent of tampon products are priced between 3.50 euro and 4 euro and that prices above 4 euro represent a minor share. Compared to tampons, we see a more balanced distribution of sanitary pads around the median price of 2.89 euro.

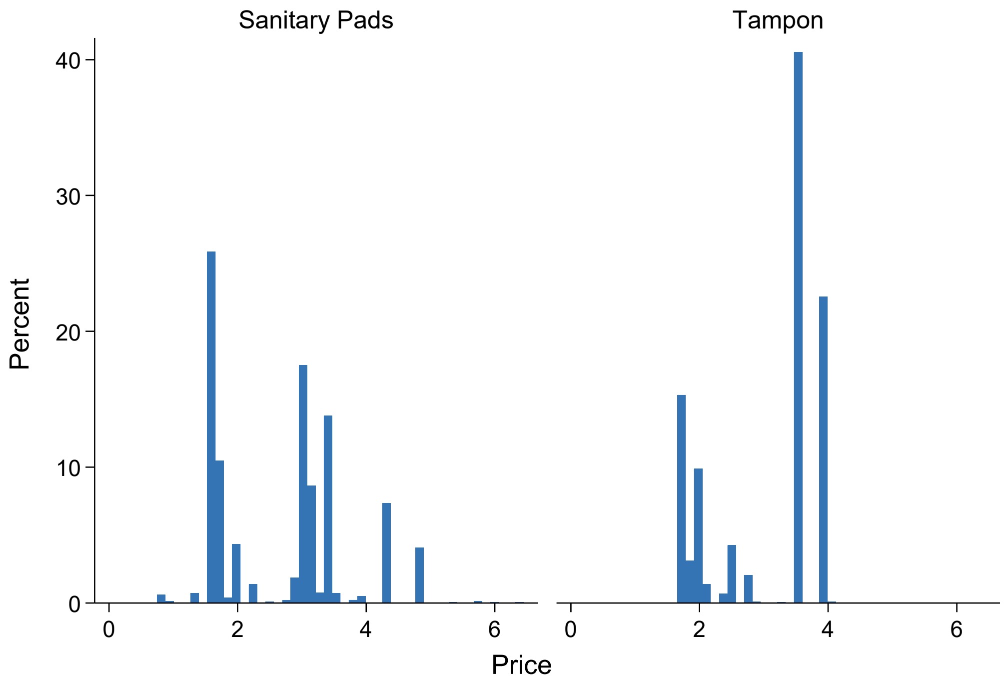


Figure A.1: This graph shows the distribution of prices for the two product types. The histogram on the left shows the percentage shares for sanitary pads. The histogram on the right shows the percentage shares for tampons.

1.B.1 Product types

We plot net prices over time by product type. Figure A.2 shows net prices normalized to $t = -11$. The average net price change for tampon products between the passage of the law and the effective date of the VAT change was about 10.5 percentage points, and the excess price reduction after the VAT change is less than one percentage point.

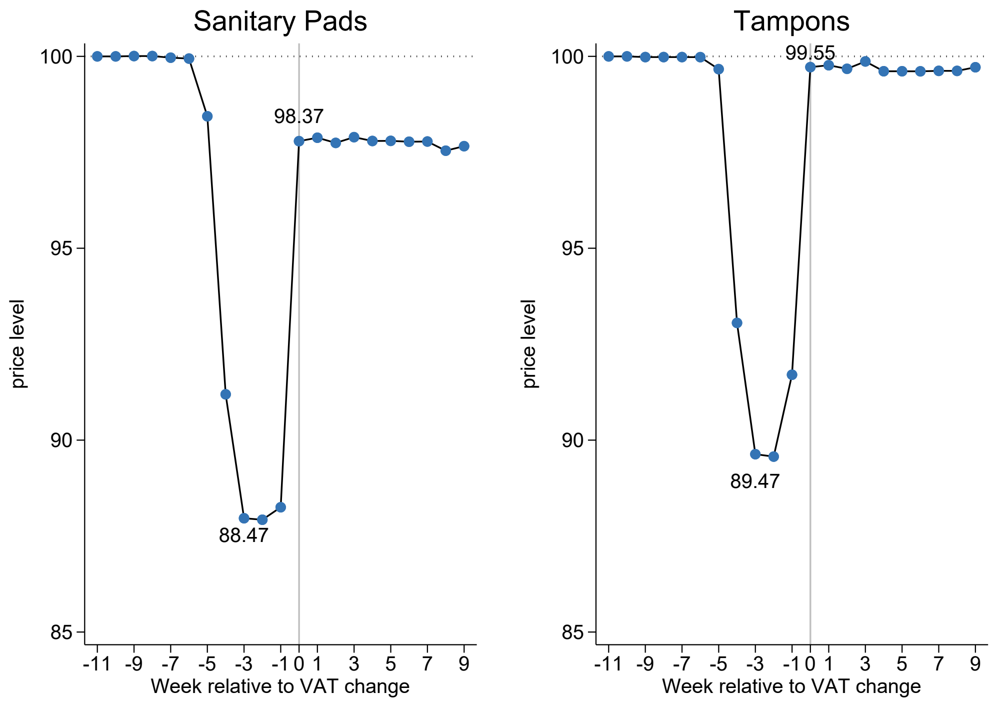


Figure A.2: This figures shows the normalized net price evolution over time separated by the product types. The graph on the left reflects the changes in net prices for sanitary pads while the graph on the right relates to tampons. $t = -5$ marks week number 48 in 2019 when the law was passed. $t = 0$ represents the first week of 2020 when the law came into effect. The price level is normalized to $t = -11$.

The net prices of sanitary pads are reduced by one percentage point more than the prices of tampons. After the reform takes effect, sanitary pad net prices remain about one percentage point below tampon prices. The temporal net price reductions show the retailer's lost profits due to an actual net price reduction. Comparing the two types, the net price decrease in the *between* phase ($t = -5$ to $t = -1$) was more restrained for tampons. After the VAT change (from $t = 0$), the figure shows that net prices were stable for both product types.

1.B.2 Store types

We plot net prices over time by store type. Figure A.3 shows net prices normalized to $t = -11$. The average net price change at drugstores between the passage of the law and the effective date of the VAT change was about 12 percentage points, and the excess price reduction after the VAT change is about two percentage points.

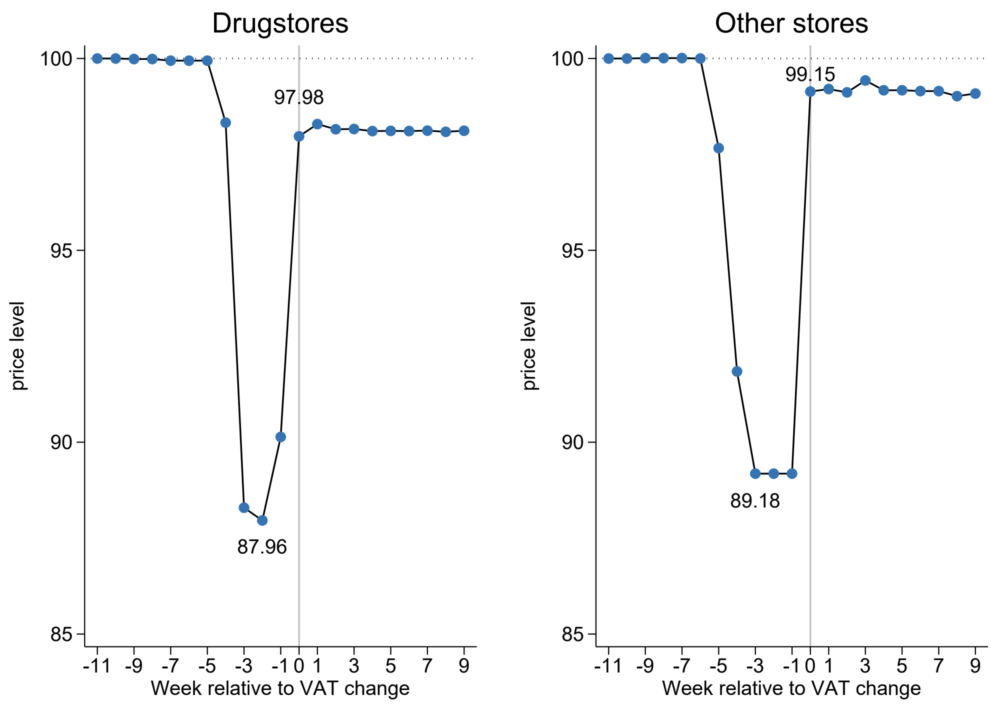


Figure A.3: This figures shows the normalized net price evolution over time separated by the store types. The graph on the left reflects the changes in net prices for drugstores while the graph on the right relates to other stores. $t = -5$ marks week number 48 in 2019 when the law was passed. $t = 0$ represents the first week of 2020 when the law came into effect. The price level is normalized to $t = -11$.

The net price reduction for grocery stores is more than one percentage point lower than for drugstores. After the VAT change, the net prices of other stores remain one percentage point below drugstore prices.

1.B.3 Competition

We plot net prices over time by competition. Figure A.4 shows net prices normalized to $t = -11$. The average change without competition between the passage of the law and the effective date of the VAT change was about 10.5 percentage points. The net price reduction after the VAT change takes effect varies between an excess pass-through of 0.5 percentage points and less than full pass-through.

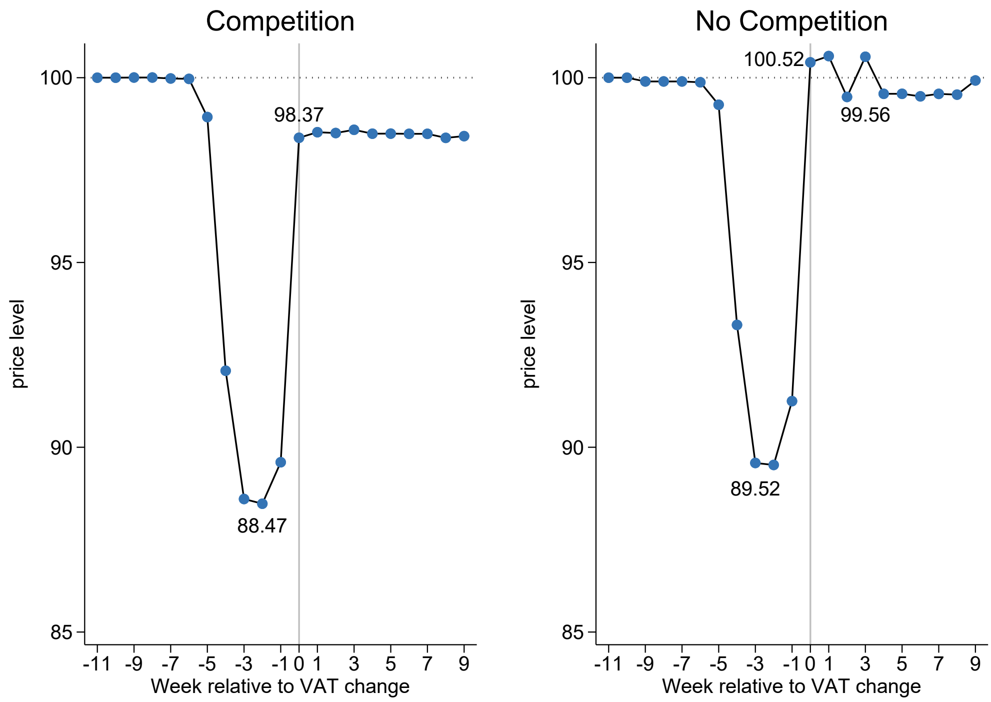


Figure A.4: This figure shows the normalized net price evolution over time, separated by competition. The graph on the left reflects the changes in net prices for products with competition, while the graph on the right refers to products without competition. $t = -5$ marks week number 48 in 2019, when the law was passed. $t = 0$ represents the first week of 2020, when the law went into effect. The price level is normalized to $t = -11$.

Net prices for products with competition are reduced by about one percentage point more than for products without competition. After the reform, net prices for products with competition remain about 1.6 percentage points below the previous level.

1.C Supplementary regression results

1.C.1 Robustness

1.C.1.1 Number of retailers

We observe in total 10 different retailers on a largely weekly basis in our data. However, one of these 10 retailers is only observed in 2019w42 and 2020w4 which means there is no observation in the weeks declared as 'between' periods. Table A.3 compares the results for 10 retailers and the results for the nine retailers when the retailer with fewer observations over time is excluded. We compare the specifications in the baseline regression and interacted with the product type dummy variable. In both regressions, the baseline with or without retailer D and when the product type is included, we see very little difference between the data sets. Therefore, we decide to also include retailer D in the analyzed data.

Table A.3: Result comparison with and without retailer D.

	10 Retailer		9 Retailer	
	=including D	+ product type	=excluding D	+ product type
between	-0.114***	-0.124***	-0.114***	-0.124***
	(0.000562)	(0.000741)	(0.000580)	(0.000646)
after	-0.119***	-0.127***	-0.119***	-0.127***
	(0.000350)	(0.000548)	(0.000327)	(0.000519)
between × tampon		0.0250***		0.0249***
		(0.00122)		(0.00112)
after × tampon		0.0181***		0.0180***
		(0.000602)		(0.000604)
Constant	1.031***	1.031***	1.032***	1.032***
	(0.00273)	(0.00271)	(0.00268)	(0.00268)
Retailer-Product FE	✓	✓	✓	✓
R^2	.9978904	.9980533	.9979253	.9980869
Observations	17015	17015	16921	16921

Notes: Bootstrapped standard errors in parentheses. This table shows the differences of the baseline estimation and the estimation including the product type dummy (× tampon, × *tampon* presents the difference for tampons) with and without retailer D. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.1.2 Handling of week number 1

The turning of the year from 2019 to 2020 was a Tuesday to Wednesday. The official WN 1 in 2020 therefore started on Monday, December 30, 2019, which means WN 1 in the data set contains both the days when the applied VAT was still 19 percent and the days when the new VAT of 7 percent was applied. As a robustness check, we change WN 1 in 2020 to the 'between' period in the regressions presented in columns (3) and (4) of Table A.4. In comparison to columns (1) and (2) in which WN 1 is specified as ‘after’, the observed effect changes accordingly. We compare the specifications in the baseline regression and interacted with the product type dummy variable. We choose the specification when WN 1 is included in the ’after’ period as the majority of days in WN 1 belong to the year 2020 when the new VAT rate is applied.

Table A.4: Result comparison switching 2020w1 from ‘after’ to ‘between’.

	Baseline		Robustness check	
	2020w1 in ’after’	+ product type	2020w1 in ’between’	+ product type
between	-0.114***	-0.124***	-0.115***	-0.125***
	(0.000573)	(0.000648)	(0.000504)	(0.000591)
after	-0.119***	-0.127***	-0.119***	-0.127***
	(0.000318)	(0.000470)	(0.000316)	(0.000470)
between × tampon		0.0250***		0.0239***
		(0.00115)		(0.000996)
after × tampon		0.0181***		0.0180***
		(0.000542)		(0.000544)
Constant	1.031***	1.031***	1.031***	1.031***
	(0.00274)	(0.00274)	(0.00274)	(0.00274)
Retailer-Product FE	✓	✓	✓	✓
R^2	.9978904	.9980533	.9978859	.9980481
Observations	17015	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the differences of the baseline estimation and the estimation including the product type dummy (× tampon, × *tampon* presents the difference for tampons) when 2020w1 is in ‘’after’ or in ‘between’. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.1.3 Handling of week number 48

The law was passed on November 29, a Friday in WN 48 in 2019. This means that WN 48 could also be included in the anticipation period captured by the *between* dummy variable in our regressions. The treatment of this week also affects our estimate of the *after* price reduction, as it affects the definition of the period before the law was passed. The following Table A.5 shows the comparison for the baseline regression and interacted with the product type dummy variable. As a robustness check, we change WN 48 in 2019 to the 'between' period in the regressions shown in columns (3) and (4) of Table A.5. Compared to columns (1) and (2), where WN 48 is specified as 'before', the observed effect changes accordingly. We choose the specification where WN 48 is included in the pre-reform period because the majority of days in WN 48 are before the official passage of the law. This is therefore the more cautious approach.

Table A.5: Result comparison switching 2019w48 from 'between' to 'before'.

	Baseline		Robustness check	
	2019w48 in 'before'	+ product type	2019w48 in 'between'	+ product type
between	-0.114***	-0.124***	-0.0892***	-0.0994***
	(0.000573)	(0.000648)	(0.000939)	(0.00125)
after	-0.119***	-0.127***	-0.121***	-0.129***
	(0.000318)	(0.000470)	(0.000326)	(0.000454)
between × tampon		0.0250***		0.0242***
		(0.00115)		(0.00181)
after × tampon		0.0181***		0.0200***
		(0.000542)		(0.000544)
Constant	1.031***	1.031***	1.033***	1.033***
	(0.00274)	(0.00274)	(0.00273)	(0.00273)
Retailer-Product FE	✓	✓	✓	✓
R^2	.9978904	.9980533	.9948947	.9950632
Observations	17015	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the differences of the baseline estimation and the estimation including the product type dummy (× tampon, × *tampon* presents the difference for tampons) when 2019w48 is in ''before' or in 'between'. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.1.4 Rounding and timing

The pricing strategy of "rounding to the preferred decimals" can explain most of the excess pass-through observed in the data. Table A.6 shows that the probability of using this strategy is not strongly affected by the reform. In fact, we find that a total of 98.5 percent of the observed prices end with one of the preferred second decimal places. Before the law was passed and after the reform took effect, the proportions of prices ending in 0, 5, or 9 are 97.39 percent (before) and 99.21 percent (after). We consider the significant but small difference between the two periods to be negligible for causality attribution given the size of the shares.

Table A.6: Rounding and timing.

	Rounding
	preferred second decimal
between	0.0197***
	(0.00256)
after	0.0189***
	(0.00230)
Constant	0.973***
	(0.00214)
Retailer-Product FE	✓
R^2	.1074017
Observations	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the probability of prices ending with one of the preferred second decimals (0, 5, or 9) for the *between* and *after* phase relative to the *before* period using a Linear Probability Model. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.1.5 Fixed effects

In Table A.7 we contrast our baseline regression for different fixed effects. The coefficients presented are *between* for the expected price decrease and *after* capturing the price decrease after the VAT change took effect. Column (1) presents the result without controlling for any time-invariant fixed effects. Retailer fixed effects are introduced in column (2) to control for time-constant characteristics of the observed retailers. Column (3) reports the results when time-constant product characteristics are included. Comparing these columns, we see that much of the variation within the panel can be captured when controlling for the time-constant product characteristics. In our analysis, we have controlled for retailer and product fixed effects, which accounts for variation in both the retailer and product dimensions and controls for a large portion of the time-constant variation within our panel. This regression result from our baseline regression is shown in column (5).

Table A.7: Comparison baseline for different fixed effects.

	No FE	Retailer FE	Product FE	Retailer FE & Product FE	Retailer-Product FE
	(1)	(2)	(3)	(4)	(5)
Anticipated price drop:					
between	-0.112***	-0.126***	-0.114***	-0.114***	-0.114***
	(0.00870)	(0.00820)	(0.000586)	(0.000556)	(0.000562)
Effective date:					
after	-0.127***	-0.128***	-0.119***	-0.119***	-0.119***
	(0.00682)	(0.00647)	(0.000381)	(0.000361)	(0.000350)
Constant	1.035***	1.037***	1.031***	1.031***	1.031***
	(0.00496)	(0.00474)	(0.00270)	(0.00270)	(0.00273)
Retailer FE	✗	✓	✗	✓	✗
Product FE	✗	✗	✓	✓	✗
Retailer-Product FE	✗	✗	✗	✗	✓
R^2	.0249888	.1238924	.9971255	.997395	.9978904
R^2 (within)	.0249888	.0291289	.8878652	.8967115	.9146051
Observations	17015	17015	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. This table shows the differences of the baseline estimation for different fixed effects. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.1.6 Comprehensive regression analysis

In the previous sections, we separately identified the product type, the store type, and the number of retailers offering a product as having a significant effect on the pass-through rate. Since we analyze these effects separately, questions about their joint significance arise. Table A.8 combines all these effects in column (4). We find all added effects are relevant and significant.

Table A.8: Comprehensive regression.

	Base	+ Product type	+ Drugstore	+ Competition
	(1)	(2)	(3)	(4)
between	-0.114***	-0.124***	-0.119***	-0.105***
	(0.000573)	(0.000648)	(0.000764)	(0.00197)
after	-0.119***	-0.127***	-0.118***	-0.107***
	(0.000318)	(0.000470)	(0.000479)	(0.000972)
between × tampon		0.0250***	0.0253***	0.0261***
		(0.00115)	(0.00115)	(0.00115)
after × tampon		0.0181***	0.0185***	0.0192***
		(0.000542)	(0.000541)	(0.000561)
between × drugstore			-0.0102***	-0.00869***
			(0.00104)	(0.00111)
after × drugstore			-0.0149***	-0.0135***
			(0.000598)	(0.000611)
between × competition				-0.0159***
				(0.00223)
after × competition				-0.0136***
				(0.00112)
Constant	1.031***	1.031***	1.031***	1.031***
	(0.00274)	(0.00274)	(0.00274)	(0.00274)
Retailer-Product FE	✓	✓	✓	✓
R^2	.9978904	.9980533	.9981332	.9981552
R^2 (within)	.9146051	.9211977	.9244321	.9253225
Observations	17015	17015	17015	17015

Notes: Bootstrapped standard errors in parentheses. *between* and *after* show the price reduction as in the baseline regression; × *tampon* adds the product type differentiation into the regression; × *competition* adds the competition effect to the regression; × *drugstore* compares for drugstores and other stores. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.2 Rounding down

In comparison to section 1.5.4 we now condition on rounding down to the same second decimals. Table A.9 provides descriptive insights into the competitive segments. The regression results presented in Table A.10 show the impact of store type, product type, and retailer competition for the same condition. We restricted the sample to those with more than a full pass-through, allowing us to identify differences in competitive segments based on the rounding strategy.

Table A.9: Rounding and heterogeneities.

(1) Product type × Rounding			(2) Store type × Rounding			(3) Competition × Rounding		
	Freq.	Percent		Freq.	Percent		Freq.	Percent
Sanitary pads			*No drugstore*			*No competition*		
Rounding=0	57	10.86	Rounding=0	123	29.78	Rounding=0	14	19.72
Rounding=1	468	89.14	Rounding=1	290	70.22	Rounding=1	57	80.28
Total	525	100	Total	413	100	Total	71	100
Tampons			*Drugstore*			*Competition*		
Rounding=0	68	17.62	Rounding=0	2	0.40	Rounding=0	111	13.21
Rounding=1	318	82.38	Rounding=1	496	99.60	Rounding=1	729	86.79
Total	386	100	Total	498	54.67	Total	840	100
Total			Total			Total		
Sanitary pads	525	57.63	*No drugstore*	413	45.33	*No competition*	71	7.794
Tampons	386	42.37	*Drugstore*	498	54.67	*Competition*	840	92.21
Total	911	100	Total	911	100	Total	911	100

Notes: This table shows descriptive insights into the heterogeneities in the 'Total' row and descriptive insights within the competitive segments conditioned on whether retailers consistently round down.

Column (1) of Table A.10 contains the product type heterogeneity, while Table A.9 column (1) shows the number of observations for the two product types differentiated by rounding. We observe that relatively more sanitary pad products round to the same decimal place than tampon products. Similarly, the regression results show a highly significant higher price reduction for sanitary pads than for tampons when both round $after$ the VAT change.[34] This finding is consistent with the results from section 1.4.2.1, which also show a larger reduction for sanitary pads.

[34] We estimate the difference by adding $after \times tampon$ and $after \times rounding \times tampon$. The estimated difference is approximately 2.3 percentage points after the VAT change.

Table A.10: Rounding with constant second decimal given excess pass-through ×

	+ Product Type	+Store type	+Competition
	ln(price)	ln(price)	ln(price)
between	-0.102***	-0.111***	-0.104***
	(0.00127)	(0.000662)	(0.00146)
after	-0.105***	-0.112***	-0.102***
	(0.000882)	(0.000520)	(0.00153)
between × rounding	-0.0258***	0.00296**	-0.00333
	(0.00144)	(0.000977)	(0.00256)
after × rounding	-0.0250***	0.000646	-0.00814***
	(0.00104)	(0.000658)	(0.00196)
between × tampon	-0.0143***		
	(0.00155)		
after × tampon	-0.0118***		
	(0.000912)		
between × rounding × tampon	0.0442***		
	(0.00211)		
after × rounding × tampon	0.0342***		
	(0.00113)		
between × drugstore		0.0548	
		(0.0410)	
after × drugstore		0.00296***	
		(0.000520)	
between × rounding × drugstore		-0.0691	
		(0.0411)	
after × rounding × drugstore		-0.0189***	
		(0.000837)	
between × competition			-0.00735***
			(0.00150)
after × competition			-0.0115***
			(0.00149)
between × rounding × competition			-0.00316
			(0.00259)
after × rounding × competition			-0.00135
			(0.00199)
Constant	1.048***	1.048***	1.048***
	(0.00275)	(0.00281)	(0.00275)
Retailer-Product FE	✓	✓	✓
R^2	.9982832	.998155	.9980903
Observations	16083	16083	16083

Notes: Bootstrapped standard errors in parentheses. This table shows the different heterogeneities interacted with a dummy identifying prices with rounding to a consistent second decimal (× rounding) given excess pass-through. In column (1) *between* and $after$ show the price reduction for sanitary pads; × *tampon* presents the difference for tampons. In column (2) *between* and $after$ show the price reduction for other stores; × *drugstore* presents the difference for drugstores. In column (3) *between* and $after$ show the price reduction without competition; × *competition* presents the difference with competition. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The second column of Table A.10 presents the store-type heterogeneity. Nearly all drugstore products are rounded to the same second decimal while 30 percent of other store products are not. The regression results indicate that drugstores reduce their prices highly significantly more than other stores given both round.[35]

The competition heterogeneity is included in column (3) of Table A.10. Table A.9 column (3) shows the number of observations with and without competition differentiated by rounding. We can see that relatively more observations with competition rounded compared to those without competition. This is also reflected in the regression results as we see highly significantly more price reduction with competition given both round.

1.C.3 Retailer

Disaggregating the data by retailers, Table A.11 shows that all retailers except retailer D[36] reduced their prices by at least the amount or even more than the VAT reduction would suggest. Retailers C and J reduced their gross prices by more than one percentage point more than most other competitors. Also, retailer E was above the average reduction in Table A.11. Table A.12 adds a robustness check as it shows that the results are largely the same for all retailers.

Table A.13 also shows differences in the pricing reactions of the retailers concerning the LMB and other brands. Retailer C reduced the prices, on average, by 1.84 percentage points more if the product was from the LMB while retailers A and B reduced other brands' products more than the LMB products.

[35] The estimated difference is around 1.3 percentage points after the VAT change.

[36] Retailer D is only observed before and after the VAT change, but not in our 'between' period.

Table A.11: Price effects for different retailers using retailer-product fixed effects.

	Retailer									
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
between	-0.109***	-0.117***	-0.116***		-0.118***	-0.0856***	-0.0856***	-0.0863***	-0.116***	-0.133***
	(0.000428)	(0.000259)	(0.00188)		(0.00249)	(0.00760)	(0.00760)	(0.0103)	(0.000268)	(0.00247)
after	-0.109***	-0.117***	-0.131***	-0.0955***	-0.123***	-0.114***	-0.114***	-0.115***	-0.116***	-0.134***
	(0.000406)	(0.000227)	(0.000838)	(0.0134)	(0.00158)	(0.000967)	(0.000967)	(0.00124)	(0.000277)	(0.00197)
Constant	0.997***	1.026***	0.937***	0.921***	0.936***	1.314***	1.314***	1.306***	1.299***	1.170***
	(0.00586)	(0.00534)	(0.00498)	(0.0526)	(0.0149)	(0.00485)	(0.00485)	(0.00502)	(0.00173)	(0.0124)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.9992449	.9997805	.9949568	.9930878	.9979056	.9421968	.9421968	.9420741	.9982031	.9971784
Observations	5010	3639	4423	94	890	189	189	168	1638	775

Notes: Bootstrapped standard errors in parentheses. This table shows the heterogeneity of price changes between retailers both in the anticipation period and after the actual VAT change. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. Retailers are anonymized using letters A to J. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.12: Price effects for different retailers using retailer-product fixed effects.

	Retailer	
	(1)	
between	-0.109***	(0.000508)
after	-0.109***	(0.000468)
A	0	(0)
B × between	-0.00806***	(0.000551)
C × between	-0.00719***	(0.00208)
E × between	-0.00930***	(0.00229)
F × between	0.0232**	(0.00847)
G × between	0.0232**	(0.00829)
H × between	0.0225*	(0.00985)
I × between	-0.00707***	(0.000559)
J × between	-0.0238***	(0.00250)
B × after	-0.00785***	(0.000504)
C × after	-0.0214***	(0.00112)
D × after	0.0138	(0.0119)
E × after	-0.0137***	(0.00169)
F × after	-0.00486***	(0.00113)
G × after	-0.00486***	(0.00103)
H × after	-0.00577***	(0.00113)
I × after	-0.00662***	(0.000549)
J × after	-0.0248***	(0.00211)
Constant	1.031***	(0.00273)
Retailer-Product FE	✓	
R^2	.9980462	
Observations	17015	

Notes: Bootstrapped standard errors in parentheses. This table shows the majority of retailers' price changes was significantly different both in the anticipation period and after the actual VAT change. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.13: Price effects for different retailers including LMB dummy variable using retailer-product fixed effects.

	Retailer									
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
between	-0.115***	-0.127***	-0.0898***		-0.113***	-0.0856***	-0.0856***	-0.0863***	-0.116***	-0.117***
	(0.00213)	(0.00178)	(0.00721)		(0.00547)	(0.00760)	(0.00760)	(0.0103)	(0.000268)	(0.00182)
between × LMB	0.00593**	0.0112***	-0.0285***		-0.00648					-0.0226***
	(0.00219)	(0.00178)	(0.00744)		(0.00541)					(0.00384)
after	-0.115***	-0.130***	-0.113***	-0.0690	-0.124***	-0.114***	-0.114***	-0.115***	-0.116***	-0.122***
	(0.00207)	(0.00131)	(0.00241)	(0.0500)	(0.00324)	(0.000967)	(0.000967)	(0.00124)	(0.000277)	(0.00159)
after × LMB	0.00546*	0.0140***	-0.0189***	-0.0355	0.00174					-0.0177***
	(0.00215)	(0.00134)	(0.00254)	(0.0501)	(0.00351)					(0.00348)
Constant	0.997***	1.026***	0.937***	0.921***	0.936***	1.314***	1.314***	1.306***	1.299***	1.170***
	(0.00586)	(0.00534)	(0.00498)	(0.0530)	(0.0149)	(0.00485)	(0.00485)	(0.00502)	(0.00173)	(0.0123)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.9992461	.999808	.9950227	.9934743	.9979153	.9421968	.9421968	.9420741	.9982031	.9973022
Observations	5010	3639	4423	94	890	189	189	168	1638	775

Notes: Bootstrapped standard errors in parentheses. This table shows the heterogeneity of price changes between retailers both in the anticipation period and after the actual VAT change differentiated by LMB and other brands. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. × LMB presents the difference for the leading manufacturer brand (LMB). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.4 Number of retailers

Table A.14 presents the results when we take the number of retailers (denoted by #) stocking a same given product into account. Because the data set does not comprise the complete supply-side, this number has to be interpreted as a proxy for the true number of retailers.

The first column in Table A.14 upper part shows the average relative price reduction for products offered at just one retailer in the data set. The relative price reduction of 10.07 percent is not significantly different from the benchmark of 10.084 percent.[37] The other columns of Table A.14 show that if there is at least one competitor also stocking the same product, the VAT reduction is stronger. The sum of coefficients for *between* and *after* is significantly different from the benchmark of 10.084 percent.

Table A.15 presents the relative price effects for different numbers of retailers offering the product with a type dummy variable. The positive sign of the *between* × *tampon* coefficients for two or more retailers support the finding of stronger effects for sanitary pads from section 1.4.2.1.

[37] The benchmark of a 10.084 percent price reduction corresponds to a regression coefficient of -0.1063.

Table A.14: Price effects by number of retailers offering the product.

	Retailer								
	(#1)	(#2)	(#3)	(#4)	(#5)	(#6)	(#7)	(#9)	(#10)
between	-0.101***	-0.114***	-0.122***	-0.113***	-0.132***	-0.123***	-0.0959***	-0.102***	-0.0977***
	(0.00207)	(0.00108)	(0.00143)	(0.000784)	(0.00312)	(0.00205)	(0.00384)	(0.00421)	(0.00318)
after	-0.106***	-0.116***	-0.127***	-0.116***	-0.136***	-0.128***	-0.106***	-0.117***	-0.113***
	(0.00108)	(0.000866)	(0.000898)	(0.000276)	(0.00231)	(0.00122)	(0.00140)	(0.000654)	(0.000527)
Constant	1.293***	0.760***	0.752***	1.241***	0.839***	0.938***	1.374***	1.280***	1.329***
	(0.0123)	(0.00654)	(0.00379)	(0.00297)	(0.00923)	(0.0107)	(0.000867)	(0.00258)	(0.00178)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.9980455	.9979896	.993637	.9973693	.9934397	.9979359	.9184223	.9667544	.9680043
Observations	1276	2400	4173	5354	519	1313	363	519	1098

Notes: Bootstrapped standard errors in parentheses. This table shows the price changes depending on the number of retailers in the anticipation period and after the actual VAT change. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.15: Price effects by number of retailers offering the product.

	Retailer								
	(#1)	(#2)	(#3)	(#4)	(#5)	(#6)	(#7)	(#9)	(#10)
between	-0.105***	-0.116***	-0.149***	-0.120***	-0.134***	-0.143***	-0.0959***	-0.105***	-0.105***
	(0.00233)	(0.00106)	(0.00167)	(0.000426)	(0.00401)	(0.00337)	(0.00384)	(0.00572)	(0.00494)
between × tampon	0.0156**	0.0182***	0.0517***	0.0193***	0.00462	0.0356***		0.00900	0.0112*
	(0.00530)	(0.00506)	(0.00224)	(0.00184)	(0.00604)	(0.00445)		(0.00925)	(0.00514)
after	-0.110***	-0.117***	-0.150***	-0.120***	-0.139***	-0.145***	-0.106***	-0.121***	-0.121***
	(0.000666)	(0.000983)	(0.00164)	(0.000369)	(0.00273)	(0.00234)	(0.00140)	(0.000835)	(0.000896)
after × tampon	0.0184***	0.00781***	0.0423***	0.0105***	0.00991*	0.0314***		0.0125***	0.0130***
	(0.00351)	(0.00183)	(0.00166)	(0.000448)	(0.00447)	(0.00267)		(0.00120)	(0.000964)
Constant	1.292***	0.760***	0.752***	1.241***	0.839***	0.938***	1.374***	1.280***	1.329***
	(0.0123)	(0.00656)	(0.00376)	(0.00298)	(0.00923)	(0.0108)	(0.000867)	(0.00262)	(0.00179)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.9981367	.9980275	.9951037	.9976403	.993524	.9982406	.9184223	.9676533	.9690486
Observations	1276	2400	4173	5354	519	1313	363	519	1098

Notes: Bootstrapped standard errors in parentheses. This table shows the price changes depending on the number of retailers in the anticipation period and after the actual VAT change differentiated by product type. *between* captures the price reduction due to the anticipatory effect; *after* shows the price reduction after the VAT change. × *tampon* presents the difference for tampons. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.C.5 Size and packsize

There are various sizes and packages for both product types. The most frequently offered tampon size is "normal" followed by the size "super," then the less often supplied "super plus" and with some more distance "mini." The normal size is mostly sold in packs of 16, as is the super size. Around 35 percent of all packages stocked include 16 tampons. The share of normal size tampons in a pack of 16 is 14.5 percent among all tampon products in our data set and this is the most frequently offered tampon package.

In the sanitary pads segment most of the products have "wings" and are followed by the "normal" sized sanitary pads. Whereas the sanitary pads with wings are mainly offered in a pack of 18 the normal size sanitary pads are sold in a 14 pack. Although the differentiation within the technologies is limited, there is a rich variety of size and package size combinations.

We analyze the pass-through also with respect to different products within the two product groups. In Table A.16 upper part, columns (2) and (3) show that the relative price reductions for the two sanitary pad types offered most in the data are smaller than the reduction at the aggregated level, but they exceed the benchmark of 10.084 percent.

Columns (6) and (7) show the two most frequently offered combinations of size and package size for sanitary pads. With relative price reductions of 11.25 percent and 12.18 percent they are also significantly above the benchmark (for (6) p-value< 0.001 and for (7) p-value< 0.001).

For the tampon technology, Table A.16 lower part shows a quite homogeneous price reduction for all types and also for the two most frequently offered combinations of size and package size.

Table A.16: Price effects by size – sanitary pads and tampons.

	Sanitary pads						
	(all types)	(wings)	(long)	(night)	(normal)	(wings#18)	(normal#14)
between	-0.124***	-0.122***	-0.131***	-0.129***	-0.121***	-0.115***	-0.129***
	(0.000657)	(0.00107)	(0.00161)	(0.00178)	(0.00111)	(0.00231)	(0.00242)
after	-0.127***	-0.123***	-0.133***	-0.133***	-0.124***	-0.119***	-0.130***
	(0.000501)	(0.000969)	(0.00120)	(0.00134)	(0.000723)	(0.000557)	(0.00231)
Constant	0.949***	0.903***	0.969***	1.012***	0.945***	1.237***	0.565***
	(0.00369)	(0.00659)	(0.00873)	(0.00885)	(0.00883)	(0.000483)	(0.00241)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓
R^2	.9977463	.9976731	.9969444	.9965979	.9986787	.9678667	.9051015
Observations	9895	3112	1915	1744	2594	656	571

	Tampons						
	(all types)	(mini)	(normal)	(super)	(superplus)	(normal#16)	(super#16)
between	-0.0994***	-0.0973***	-0.0959***	-0.104***	-0.0987***	-0.0943***	-0.0943***
	(0.00109)	(0.00309)	(0.00191)	(0.00130)	(0.00200)	(0.00300)	(0.00300)
after	-0.109***	-0.108***	-0.109***	-0.109***	-0.108***	-0.108***	-0.108***
	(0.000313)	(0.000612)	(0.000507)	(0.000635)	(0.000595)	(0.000883)	(0.000883)
Constant	1.146***	1.073***	1.055***	1.061***	1.309***	0.737***	0.737***
	(0.00412)	(0.0127)	(0.00693)	(0.00764)	(0.00380)	(0.00476)	(0.00476)
Retailer-Product FE	✓	✓	✓	✓	✓	✓	✓
R^2	.9981577	.9986113	.9980311	.9984727	.9914836	.989437	.989437
Observations	7120	865	2184	1863	1295	1097	1097

Notes: Bootstrapped standard errors in parentheses. This table shows the price changes of different sizes for sanitary pads (upper part) and tampons (lower part) including the most frequent combinations of size#packsize. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Chapter 2

Occupational Licensing and Quality in a Market for Experience Goods: Experimental Evidence

Co-authored with Alexander Rasch, Nicolas Fugger, and Carina Fugger

2.1 Introduction

The regulation of occupations is a widespread phenomenon in jurisdictions around the world and is becoming increasingly popular in many major economies, as documented in a recent review by Kleiner (2017). Regulations typically take a variety of forms (e.g., restrictions on entry, quality, pricing, capacity). One common approach is occupational licensing (for an overview see, for example, Kleiner, 2000). Occupational licensing is often found in very important industries and sectors (e.g. education, liberal professions, health). The justifications for this type of barrier to entry, which requires sellers or providers of certain services to have specific training, are manifold (including safety and competition issues). In particular, claims of minimum quality protection (Kleiner and Krueger, 2010) and – related to this – benefits for consumers are put forward. The need for consumer protection is mainly based on the fact that consumers (or buyers in general), unlike sellers who know the quality of the product, cannot determine the quality of the product at the time of purchase because they lack the necessary experience, knowledge, training, etc. (information asymmetry). In contrast, the quality can be observed only after the purchase, i.e. these products/services are so-called experience goods (Nelson, 1970, Riordan, 1986). Moreover, search costs may be prohibitively high, so that seeking different opinions or benefiting from previous consumer experiences is not an option. As a result, markets are prone to severe information failures (see, for example, Akerlof, 1970 and Leland, 1979). Occupational licensing is seen as a way to reduce uncertainty about product or service quality, because consumers can be sure that sellers are in principle able to provide a certain quality, and that this quality can be better assessed.

In competition-policy circles, the question of introducing occupational licensing is debated controversially. A prominent example is the German crafts sector.[1] Before 2004, providers of craft services were required to have a so-called master craftsman's diploma or advanced vocational certificate (*Großer Befähigungsnach-*

[1] For a characterization of the degree of information asymmetry for different services in the crafts sector, see Röber (2009).

weis/Meistertitel). This certificate was a prerequisite for entering the market in the first place (*'Meisterzwang'*). In 2004, the German government decided to reform the crafts regulation. The reform was intended to lower barriers to entry in order to increase competition and employment in the sector. To this end, the previous licensing scheme of a mandatory certificate was abolished or at least softened for a number of trades. As a consequence, obtaining a master craftsman's diploma in these deregulated trades was voluntary. One reason given by opponents of deregulation was that there were concerns about the effectiveness of an optional diploma in ensuring consumer protection. For some time, some political parties and industry associations have called for a return to the more restrictive pre-2004 occupational licensing scheme. Eventually, calls for re-licensing were heard, and as of 2020, a large number of previously de-licensed trades nowadays require occupational licenses again (*Rückvermeisterung*).[2] A similar development could be observed for the licensing requirements for barbers in Alabama, USA.[3]

Motivated by these mixed regulatory efforts to improve outcomes in markets for experience goods, we address the question of whether regulation in the form of occupational licensing (as a barrier to entry) has positive or negative effects (if any), especially with respect to consumer surplus. To this end, we conduct an experiment to analyze two different scenarios with respect to differences in information provision through licensing or mandatory qualification: An investment in qualification can either signal the exact quality of a product/service (observability/transparency), or it can signal that a firm is in principle capable of providing high quality (ability).[4] The market is served by two sellers who sequentially choose whether to invest to obtain a license; thereafter, the sellers choose a quality and a price for the product offered. In both scenarios, we perform two treatments: a treatment in which both sellers

[2]Haucap and Rasch (2019) discusses various aspects that argue for regulation in the German crafts sector. A critical assessment of regulation in this sector can be found in Recker (2018).

[3]Timmons and Thornton (2019) provides an analysis of the effects on barbers' profits after de-licensing and re-licensing. For an earlier account of the positive effects of licensing on barber profits, see Timmons and Thornton (2010).

[4]We acknowledge that in reality mandatory qualification can be expected to affect both aspects. We distinguish between these two extremes in order to trace the effects of regulation on each dimension.

can always offer a product regardless of their investment decision, and a second treatment in which investment is required to offer a product. The second treatment is designed to reflect the case of occupational licensing as a barrier to entry.

Our experimental results show that, partially in line with our theoretical predictions, sellers invest significantly more in both scenarios when investment is a prerequisite for entry. However, we find more investment than predicted in both treatments in the scenario where buyers never observe quality. At the same time, in contrast to the prediction, we do not see a significant change in quality across treatments in this scenario. When investment leads to observability of quality prior to potential trade, we observe significantly higher quality. Consumer surplus also tends to increase in this scenario. Our results suggest that a barrier to entry in the form of occupational licensing leads to higher quality if the investment leads to more transparency with respect to the quality offered. This is not to the detriment of consumers, as consumer surplus remains unchanged and tends to be even higher. On the other hand, if the quality of the product remains unknown to the buyers, such an entry barrier leads to a lower consumer surplus.

Our paper contributes to the experimental literature analyzing markets characterized by asymmetric information. The paper most closely related to ours is the analysis by Henze et al. (2015). The authors consider an experience good market that is similar to ours, but they are interested in the role of information disclosure. To do so, they vary the degree to which consumers are informed about quality. The authors find that, contrary to what is predicted by the theory, firms do not differentiate quality under full information, but tend to offer products of similar high quality, leading to more intense price competition than expected. When consumers have no information, there is a "lemons" outcome with low quality, but prices are substantially above marginal cost.[5] When information disclosure is of intermediate degree, quality is significantly higher than in the treatment without information,

[5] Such an outcome is also observed in the earlier contributions by Lynch et al. (1986) and Holt and Sherman (1990), where – in contrast to Henze et al. (2015) and to our setup – vertical differentiation is not possible due to buyer homogeneity.

and there is evidence that prices become better predictors of quality. The authors conclude that information disclosure is a more effective means of increasing welfare and consumer surplus than theory predicts. We essentially take their treatment without information and add an investment stage that can have an information disclosure effect; thus, our main contribution is to endogenize ability and information levels.

Huck et al. (2012, 2016) study an experience good market in the form of a binary-choice trust game, where they analyze different matching procedures. In their treatments, buyers are either matched with a single seller or buyers can choose a seller from the set of sellers. In contrast, in the present setup, two sellers always compete for buyers (as in Henze et al., 2015). Furthermore, reputation building, which is the focus of Huck et al. (2012, 2016), is not possible in our experiment due to random matching in each period.[6] Our focus is on the effects of information disclosure and entry.[7]

In her experimental study of a market with costly buyer search and differentiated sellers, Mago (2010) investigates the effect of buyer information on price or product characteristics. The author finds that information about product characteristics can raise market prices, implying that the presence of informed buyers creates negative externalities for uninformed buyers. In our experiment, the level of information is the same for all buyers.

The rest of this paper is organized as follows. In section 2.2 we analyze the theoretical background of our study. Sections 2.3 and 2.4 describe the experimental design and procedures. Based on the theoretical background, we present our hypotheses in section 2.5. In section 2.6 we compare our experimental results with the hypotheses. Section 2.7 discusses subgames beyond the equilibrium paths. In section 2.8, we

[6]Seller reputation in a market for experience goods is also studied by Cason and Gangadharan (2002). They study this issue in the context of environmental product labeling. Introducing different treatments such as seller reputation, "cheap talk" signals, and third-party certification to address the market failure arising from incomplete information, they find that reputation sometimes increases the number of higher quality products, but that third-party certification most reliably achieves the goal.

[7]In a recent experimental study, Bochet and Siegenthaler (2021) find that bargaining can improve efficiency in markets with asymmetric information.

discuss our results. We conclude in section 2.9.

2.2 Theoretical background

In this section, we set up a theoretical framework for our experiment. We then present the equilibria in the different scenarios we are interested in.

2.2.1 Framework

We consider the following experience good market. There are two sellers, indexed by $i \in \{A, B\}$. Sellers charge a price p_i for their product. Buyers may not assess the product quality q_i at the time of purchase. Depending on sellers' investment decisions (see below), sellers may choose qualities from the set $Q = \{1, 2, \ldots, 10\}$. Sellers' linear production costs per unit sold depend on the quality offered. They face costs of q_i per unit sold. It is more costly for sellers to offer high quality than low quality. Investment costs amount to 16. There is no other fixed cost and units are produced on demand.

There are four buyers, each buying at most one unit of the product. All buyers get the same fixed benefit of $v = 20$ from the purchase. Furthermore, buyers differ in their taste for quality: A buyer k's taste for quality is measured by θ_k, where θ_k is independently drawn from a uniform distribution on the support $[\underline{\theta}, \bar{\theta}]$.[8] Whereas a buyer's taste for quality is her private information, sellers know the distribution from which it is drawn. A buyer receives a utility of

$$u_j = \begin{cases} 20 + \theta_k q_i - p_i & \text{if she buys from seller } i \\ 0 & \text{if she does not buy.} \end{cases}$$

Utility increases linearly with a buyer's taste for quality and decreases with price. Whether a buyer can assess a product's quality at the time of purchase depends on the effect of an investment (see below).

[8] In the experiment, we choose $\underline{\theta} = 1$ and $\bar{\theta} = 4$.

The timing of the game is as follows:[9]

1. Seller A chooses whether to invest.
2. Seller B observes seller A's investment decision and decides whether to invest.
3. Seller A observes seller B's investment decision, and sellers simultaneously set their quality levels.
4. Sellers observe each other's qualities and simultaneously set prices.
5. Buyers make purchasing decisions after observing prices. Depending on the investment decisions and the treatment, they may also observe quality.

In the experiment, we analyze four scenario-treatment combinations that differ in the effect that an investment has and in the requirement to invest as a prerequisite to offer a product in the first place. As a consequence, the information available to buyers differs.

Ability scenario: An investment enables a seller to choose a quality from the above-mentioned set of qualities. By contrast, the quality of a product offered by a seller who does not invest in qualification remains at the lowest level of 1. In other words, the investment in Ability makes it possible for a seller to deliver a high-quality product, but a potentially high quality is not observable for the buyers at the time of purchase (experience good). By contrast, buyers can perfectly tell the low quality of a product whose seller did not invest.

Transparency scenario: Sellers can always choose from the above set of qualities irrespective of their investment decision. If a seller decides to invest, this means that buyers can observe the quality chosen by the seller. As a consequence, if a seller invests, buyers know the product quality before a potential transaction takes place.

[9]The assumption of sequential investment decisions simplifies the equilibrium analysis without affecting the overall investment level. Moreover, it ensures that pure-strategy equilibria exist. From an application point of view, it mirrors a real-life situation in which a potential entrant observes what the industry standard with regard to investment is.

NoBarrier treatment: Sellers are free to choose whether to invest. Independent of their investment decision, they can always offer their products. This scenario mirrors a situation in which no occupational licensing is in place.

Barrier treatment: Only those sellers who invest can offer a product. As such, the scenario represents the case in which sellers must obtain occupational licensing as a prerequisite to enter the market.

In the following, we analyze combinations of investment effects and entry. Thus, we consider the following four treatments: AbilityNoBarrier (A/NB), AbilityBarrier (A/B), TransparencyNoBarrier (T/NB), and TransparencyBarrier (T/B).

2.2.2 Equilibria

In this section, we present the equilibria in the four scenario-treatment combinations. For a derivation of the equilibria, see Appendix 2.A.

AbilityNoBarrier. There exist two pure-strategy perfect Bayesian equilibria. In one of them, both sellers do not invest in qualification and, hence, provide products of the lowest quality; they face Bertrand-style competition. In the other equilibrium, one seller (seller A) invests in qualification and the invested seller chooses the maximum quality. The other seller has a quality of 1. In this equilibrium, both sellers make positive profits. Buyers' beliefs determine whether investment is a credible signal of high quality.

AbilityBarrier. There exists a unique pure-strategy subgame-perfect equilibrium in which seller A invests and seller B does not invest (i.e., does not enter the market). Moreover, seller A offers the lowest quality, which is 1. Seller A sets the monopoly price such that it is just accepted by all buyers given the low quality, which is therefore 21.[10]

[10]Although maximum differentiation yields the highest profits for both sellers, the low-quality seller (seller B) does not enter the market. The expected profit of 12 for the low-quality seller is

TRANSPARENCYNOBARRIER. There exists a unique pure-strategy perfect Bayesian equilibrium in which seller A invests and seller B does not invest. Seller A provides a high quality of 10 at a price of 28, and seller B provides a low quality of 1 at a price of 10.

TRANSPARENCYBARRIER. There exists a pure-strategy subgame-perfect equilibrium in which seller A invests and seller B does not invest. Seller A sets a quality of 10 and a monopoly price of 35. In this equilibrium, a fraction of buyers with a low valuation of quality will not buy any product, but extracting more rents from the buyers with a higher preference for quality is more profitable for seller A than serving the entire market.

2.3 Experimental design

The experiments were conducted online in sessions of 12 subjects, four of whom took the role of a seller and eight of whom took the role of a buyer. These roles remained fixed throughout the experiment. In each period, the 12 subjects were randomly assigned to one of two separate markets, each market consisting of two sellers and four buyers.

Subjects play the game as in section 2.2. There are 20 periods in total. Participants know the number of periods from the instructions. In each period, there is random rematching of the 12 subjects into the two markets. Also, buyers receive an individual valuation factor for quality ($\theta_k \sim \mathcal{U}[1,4]$) in each period. This factor is their private information.

There are four decisions in each period. First, sellers sequentially decide whether to invest in qualification. The choice is either to invest or not to invest. Second, sellers choose a quality level. Sellers' choices at this stage depend on the investment decision and the treatment, provided they offer a product. When sellers can choose, the action set for the quality choice is $\{1,2,\ldots,10\}$. Third, sellers choose a price from

less than the fixed cost of entry of 16.

Table 2.1: Treatment overview.

	Ability: Invested	Ability: Not Invested	Transparency: Invested	Transparency: Not Invested
NoBarrier	$Q \in \{1, 2, \ldots, 10\}$ $Q \rightarrow$ not visible entry ✓	$Q \in \{1\}$ $Q \rightarrow$ "visible" entry ✓	$Q \in \{1, 2, \ldots, 10\}$ $Q \rightarrow$ visible entry ✓	$Q \in \{1, 2, \ldots, 10\}$ $Q \rightarrow$ not visible entry ✓
Barrier	$Q \in [1, 10]$ $Q \rightarrow$ not visible entry ✓	– – entry ✗	$Q \in [1, 10]$ $Q \rightarrow$ visible entry ✓	– – entry ✗

Notes: This table shows the characteristics of each scenario and treatment. For each combination of scenario (columns) and treatment (rows), the consequences of investing in qualification (left) and not investing (right) are compared. Q is the product quality that sellers choose after making the decision to invest in qualification. This product quality Q can be either visible or invisible to buyers when they decide to trade. If sellers do not enter the market (entry ✗), they cannot offer a product and therefore do not decide on a quality or set a price.

the set {0,1,2,...,60}. Fourth, buyers observe sellers' investment decisions, prices, and product quality – depending on the scenario and treatment. Buyers then decide whether to buy and, if so, which product to buy.

Table 2.1 summarizes the treatment design. For both scenarios, Ability and Transparency, we vary the consequences of not investing in qualification. While the setting for sellers who invest in qualification is the same in each of our two scenarios, the consequences of not investing differ. These differences are exploited in the analysis of our experimental results.

2.4 Procedures

The experiment was programmed in oTree (Chen et al., 2016) and conducted online. In addition to the oTree interface, there was a parallel web-conference call with the participants and the experimenter. In this call, participants could privately ask questions for clarification and, if required, receive technical assistance. Our procedures are similar to those in Danz et al. (2021). The webcam-on protocol has proven to be useful: Li et al. (2021) argue that the results are comparable to offline lab experiments, but find that the additional webcam-on setting is less noisy. Importantly, participants in our experiment could not see each others' videos, which maintains privacy and anonymity.

In the beginning of a session, participants read the instructions. The instructions were accessible to the participants during the whole course of the experiment. After reading the instructions, the participants had to answer several control questions. When a control question was correctly answered, participants were shown additional explanations for the question. Wrong answers were highlighted and participants were given another try.

After two non-incentivized trial periods, participants played 20 incentivized periods of the experiment. All periods were payed at the end of the experiment. After the last period, a short survey about risk preferences and personal characteristics followed. The survey questions were non-incentivized.

The experiments were conducted in February and March 2022. Participants were recruited using ORSEE (Greiner, 2015) from the subject pool of the Cologne Laboratory for Economic Research (CLER) at the University of Cologne. A total of 288 participants took part. For each of the four treatments, we conducted six sessions with 12 participants. Every subject participated in only one treatment. Sessions lasted 60 to 90 minutes. We used an experimental currency unit (ECU) at an exchange rate of two Euro cents per one ECU. Participants received an initial capital of 350 ECU (7.00 Euro) at the start of the experiment. Subjects earned on average 16.37 Euro, including a show-up fee of 2.50 Euro.

2.5 Hypotheses

The following hypotheses are derived from the model in section 2.2 and tested in section 2.6. The pre-registration of these hypotheses can be found on OSF

Registries.[11] We derive our hypotheses for each of the two scenarios (Ability and Transparency), addressing the expected effects of a market entry barrier. To this end, we pairwise compare the predictions for the NoBarrier and Barrier treatments based on the equilibria from section 2.2. For each scenario, we group our hypotheses into three bundles: first, investment; second, quality and prices; and third, welfare implications.

2.5.1 Ability scenario

2.5.1.1 Investment

In the Ability scenario, the total number of investments in qualification is either 0 or 1 in the NoBarrier treatment and 1 for the Barrier treatment. We can thus state the following hypothesis:

Hypothesis 1 *There is less investment in qualification in AbilityNoBarrier than in AbilityBarrier.*

2.5.1.2 Quality and prices

In the NoBarrier treatment, the predicted quality is either 1 or 5.5 on average. For the Barrier treatment, the predicted quality is the lowest possible, which is 1. We formulate:

Hypothesis 2 *The average quality is higher in AbilityNoBarrier compared to AbilityBarrier.*

In the Barrier treatment, the market is served by a monopolist in equilibrium and the predicted average price is 21. For the NoBarrier treatment, where both sellers offer a product, the predicted average price is either 1 in the "lemons" outcome or 19 with quality differentiation. Thus, we can state the following hypothesis:

Hypothesis 3 *Average prices are lower in AbilityNoBarrier than in Ability-Barrier.*

[11] The pre-registration can be accessed using the following link: https://doi.org/10.17605/OSF.IO/PAQ56.

2.5.1.3 Welfare implications

The expected consumer surplus in the BARRIER treatment is 10. For the NOBARRIER treatment, the expected consumer surplus is either 86 for the “lemons” equilibrium or 62 for the quality differentiation.[12] We hypothesize:

Hypothesis 4 *Average consumer surplus is higher in ABILITYNOBARRIER than in ABILITYBARRIER.*

The predicted total welfare in equilibrium is 134 for the NOBARRIER treatment. For the BARRIER treatment, the predicted total welfare amount to 74.[13]

Hypothesis 5 *Average welfare is lower in ABILITYBARRIER than in ABILITYNOBARRIER.*

2.5.2 Transparency scenario

2.5.2.1 Investment

In the TRANSPARENCY scenario, a total number of one investment is predicted for both treatments. Thus, we hypothesize:

Hypothesis 6 *(Null) The investment levels are the same across TRANSPARENCYNOBARRIER and TRANSPARENCYBARRIER.*

2.5.2.2 Quality and prices

In the NOBARRIER treatment, the average quality in equilibrium is 5.5, while for the BARRIER treatment the prediction is a quality of 10 in equilibrium. For the average market qualities, we formulate the following hypothesis:

Hypothesis 7 *The average quality in TRANSPARENCYNOBARRIER is lower than in TRANSPARENCYBARRIER.*

[12] Consumer surplus is derived based on the average buyer's expected valuation of quality $\mathbb{E}[\theta] = 2.5$.

[13] Total welfare is derived based on the average buyer's expected valuation of quality $\mathbb{E}[\theta] = 2.5$.

For the NoBarrier treatment, the average price in equilibrium is 19. In the Barrier treatment, the equilibrium price is 35. Accordingly, our hypothesis reads:

Hypothesis 8 *The average prices are lower in TransparencyNoBarrier than in TransparencyBarrier.*

2.5.2.3 Welfare implications

The consumer surplus in equilibrium is 62 in the NoBarrier treatment. For the Barrier treatment the consumer surplus in equilibrium is $33.\bar{3}$.[14] As a hypothesis we formulate:

Hypothesis 9 *The average consumer surplus is higher in TransparencyNoBarrier than in TransparencyBarrier.*

The predicted total welfare in the NoBarrier treatment is 134. For the Barrier treatment, the predicted total welfare is exactly the same with 134.[15] Thus, we hypothesize:

Hypothesis 10 *(Null) Welfare is the same between TransparencyNoBarrier and TransparencyBarrier.*

We test these hypotheses in the next section.

2.6 Results

We analyze the effects of the entry barrier in the two scenarios Ability and Transparency separately. To show the effect of occupational licensing as a barrier to market entry, we compare the results for the NoBarrier treatments with those for the Barrier treatments.

We begin this section with a descriptive table and explanatory remarks in section 2.6.1. We then provide a more detailed analysis of the relevant choice variables

[14] Consumer surplus is derived based on the average buyer's expected valuation of quality $\mathbb{E}[\theta] = 2.5$.

[15] Total welfare is derived based on the average buyer's expected valuation of quality $\mathbb{E}[\theta] = 2.5$.

(*investment*, *quality*, and *price*) and market variables (*consumer surplus* and *welfare*) in section 2.6.2 for the Ability scenario and section 2.6.3 for the Transparency scenario. In these sections, we focus on our pre-registered hypotheses and test whether the experimental data matches them. In the following section 2.7, we then analyze the different subgames in more detail.

For hypothesis testing, we use panel regression models (Pooled OLS) to estimate treatment effects with and without time trends. Due to our random re-matching design, we use standard errors clustered at the session level. Unless otherwise stated, we report two-sided p-values throughout. As a further conservative robustness check, we perform non-parametric tests (ranksum tests) that can be found in Appendix 2.B.3.

2.6.1 Overview

Our main research question is whether a market entry barrier improves market outcomes in markets for experience goods. The theoretical analysis highlights that the introduction of an entry barrier crucially affects the decision to invest and subsequently qualities and prices offered in the market. Table 2.2 shows the means and standard deviations of our main variables. As expected, the realizations of the scenarios differ – especially in the quality dimension and, accordingly, in total welfare. The overall price level is low in comparison to predicted price levels, and the consumer share of total welfare is quite high in both scenarios.

We now analyze the two scenarios separately. We test whether and how the entry barrier changes the market outcomes in each scenario.[16]

[16]Importantly, we do not make qualitative statements comparing across the two scenarios because they represent two coexisting ways of interpreting the effect of occupational licensing.

Table 2.2: Descriptive table.

	Ability		Transparency	
	NoBarrier	Barrier	NoBarrier	Barrier
Investment	0.90	1.89	0.83	1.80
	(0.241)	(0.159)	(0.139)	(0.309)
Quality	2.23	1.99	4.05	7.83
	(0.702)	(0.624)	(0.931)	(1.205)
Price	12.46	15.64	14.00	18.35
	(5.574)	(3.606)	(5.271)	(4.757)
CS	62.28	38.68	76.70	88.88
	(16.59)	(17.60)	(22.83)	(22.59)
Welfare	79.08	57.78	95.07	100.25
	(3.141)	(7.376)	(8.406)	(7.096)
N	6	6	6	6

Notes: We report means and standard deviations in parentheses for the following variables: Investment: sum of sellers who invested ($\{0, 1, 2\}$). Quality: average quality offered by sellers (1–10). Price: average price set by sellers (1–60). CS: average consumer surplus in the market. Welfare: average total welfare in the market.

2.6.2 Ability scenario

In this scenario, investment in qualification enables sellers to offer higher product qualities, whereas sellers who choose not to invest can only offer the lowest possible quality. The chosen quality is not observable to buyers in this scenario when sellers have invested in their qualification. Quality is indirectly known for products from sellers who have not invested in qualification because they have no quality choices. In the following subsections, we test the pre-registered directed hypotheses 1–5 for this scenario.

2.6.2.1 Investment

Figure 2.1 illustrates the average number of investments in qualification over time across our two experimental variants. The average investments experience a substantial decline in the NoBarrier treatment and converge to values between 1.0 and

0.5, with a clear trend toward the lower bound.[17] The average investment over the last five periods is 0.55 for this treatment. The number of investments is stable just below the upper bound of 2 in the BARRIER treatment.

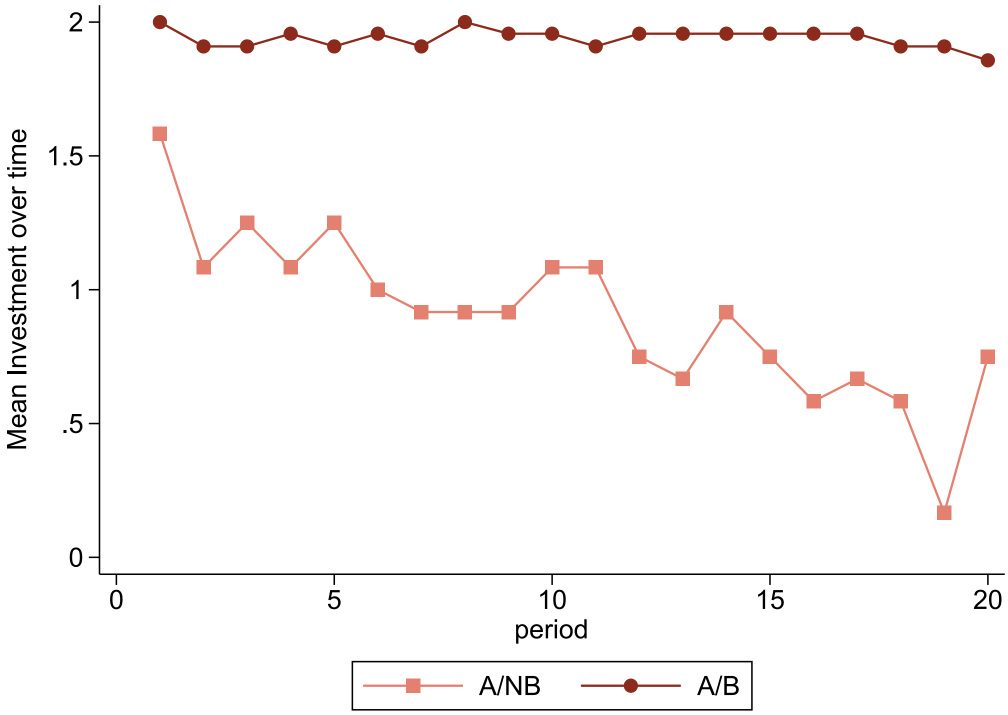


Figure 2.1: Mean number of investments in qualification in the market over time (ABILITY).

Result 2.1 *In the ABILITY scenario, the average total investment in the BARRIER treatment is significantly higher compared to the NOBARRIER treatment.*

Table 2.3 presents regressions to examine the impact of the entry barrier on investment decisions in columns 1a and 1b. As predicted, the total number of investments is significantly higher in the BARRIER treatment compared to the NOBARRIER treatment. Our tests and the analysis of the data show that there is support for our Hypothesis 1. The entry barrier significantly increases the number of investments in qualification (p-value = 0.023). The absolute treatment difference

[17] Over the 20 periods, the estimated number of investments decreases by 0.044 each period (p-value = 0.003).

is either 1 or 0.6 when we control for the time trend.

2.6.2.2 Quality

Figure 2.2 illustrates the average quality offered in the markets over time. The qualities of both treatments fluctuate around the value of 2. For the NoBarrier treatment, the time trend is significantly negative.[18] This may be related to the decreasing trend in investment because, according to the experimental design, the quality offered is equal to one if a seller does not invest. For the Barrier treatment, we find no significant time trend in the average quality offered.

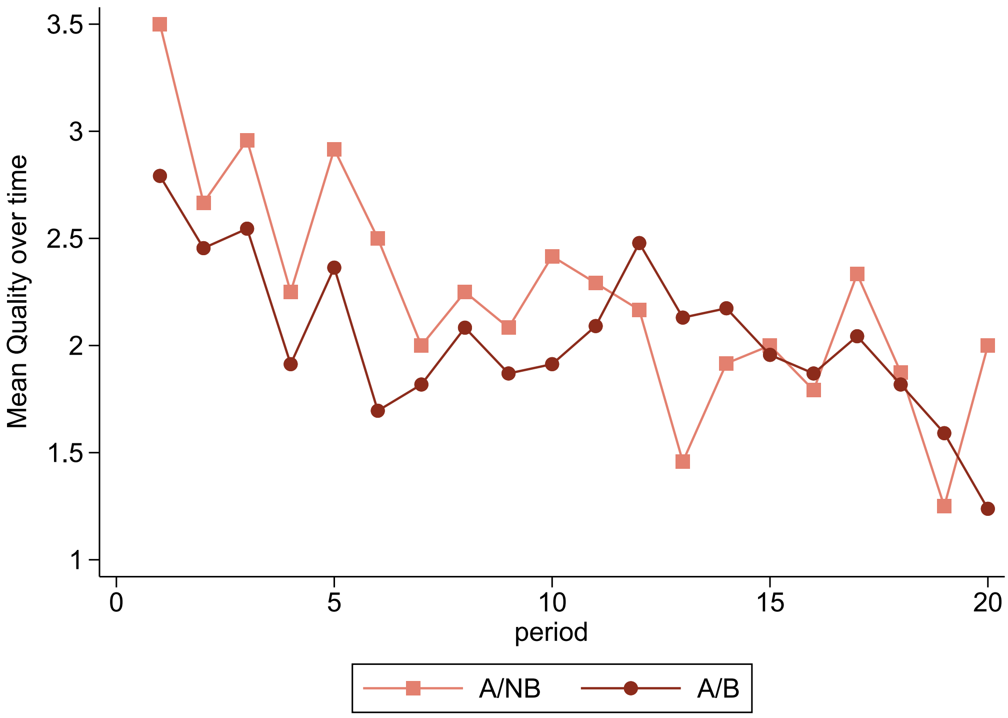


Figure 2.2: Mean quality in the market over time (Ability).

Result 2.2 *In the Ability scenario, the average qualities in the market between the Barrier treatment and the NoBarrier treatment do not differ significantly.*

Our Hypothesis 2 is that quality is higher for the Ability scenario without

[18]The average quality decreases by 0.067 (p-value = 0.006) with each period.

the regulatory barrier to entry. This prediction is not supported since we find no significant difference in qualities in Table 2.3 columns 2a and 2b.

2.6.2.3 Prices

Figure 2.3 shows the average price for the products in the market over time. For both treatments, we observe a significant downward trend over all 20 periods.[19] This trend is less pronounced for the NoBarrier treatment in the second half of the experiment.

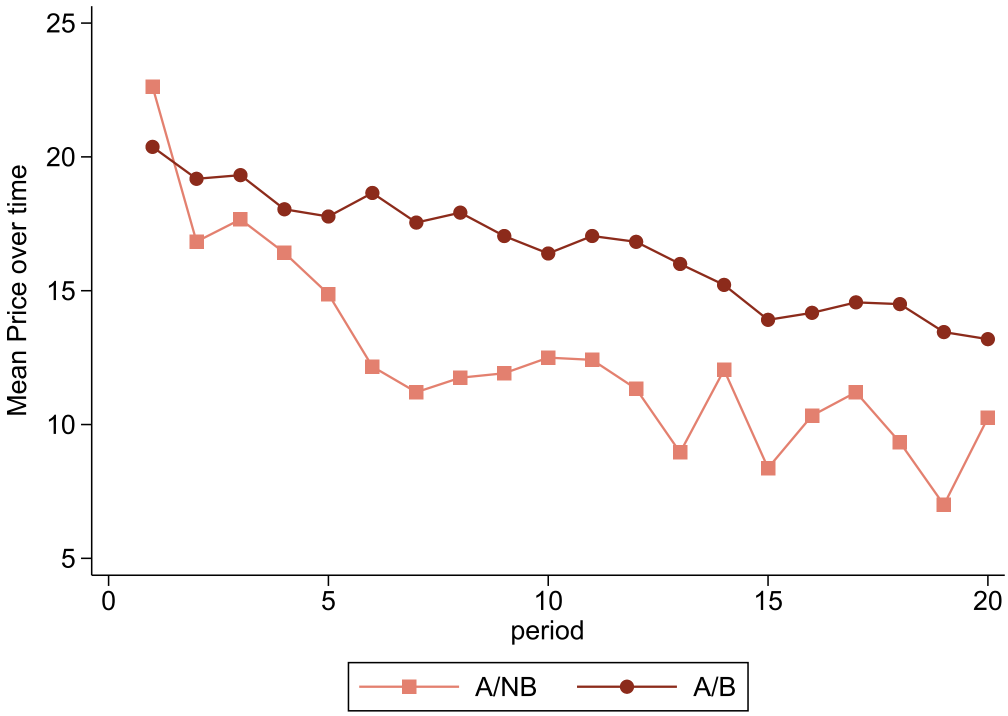


Figure 2.3: Mean price in the market over time (Ability).

Result 2.3 *In the Ability scenario, the average prices in the market between the Barrier treatment and the NoBarrier treatment do not significantly differ.*

For the Ability scenario, we hypothesize that prices are higher in the Barrier treatment compared to the NoBarrier treatment. Except for the first period,

[19] The average price decreases significantly in the NoBarrier (p-value < 0.001) and the Barrier (p-value < 0.001) treatments.

the observed averages over time follow this prediction. Nevertheless, as the results presented in Table 2.3 columns 3a and 3b show, the difference in prices between the two treatments is not significant in this scenario. This means that Hypothesis 3 is not supported by our data.

2.6.2.4 Consumer surplus

Figure 2.4 plots the average consumer surplus over time. We observe a stable and increasing consumer surplus for both treatments, which is consistent with the price decline documented in section 2.6.2.3.

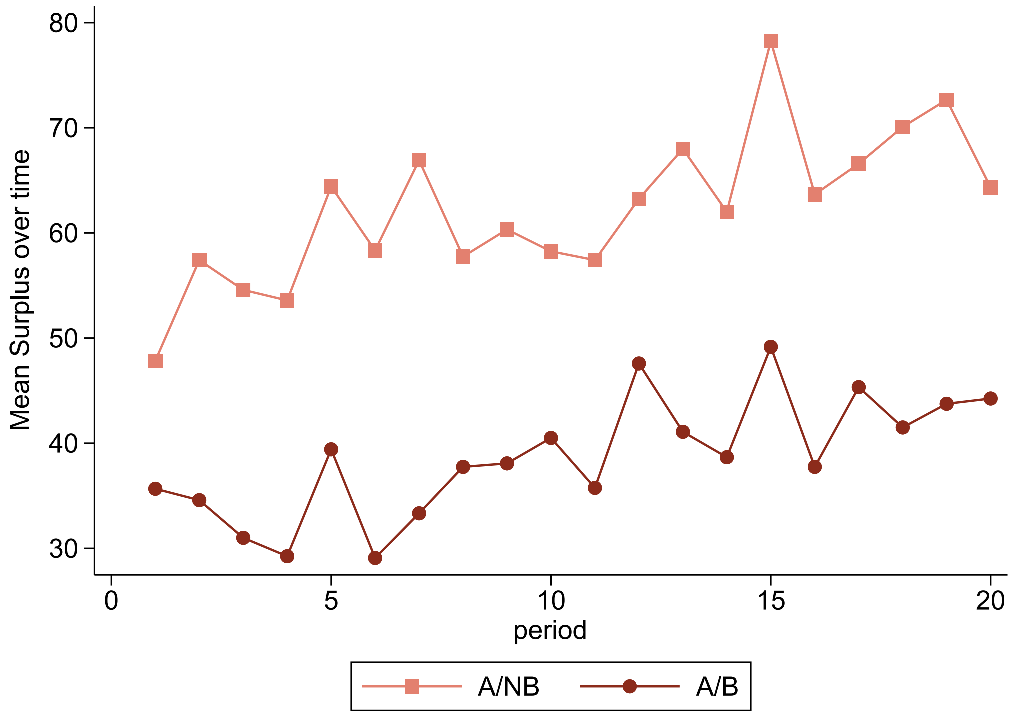


Figure 2.4: Mean consumer surplus in the market over time (Ability).

Result 2.4 *In the Ability scenario, the average consumer surplus in the market is significantly higher in the NoBarrier treatment compared to the Barrier treatment.*

The prediction is that the consumer surplus is lower with the entry barrier in the Ability scenario (Hypothesis 4). Consistent with this prediction, our results, presented in Table 2.3 columns 4a and 4b, show a significant difference between the consumer surplus in the two treatments. This rejection of the null suggests that Hypothesis 4 is supported by our data. The regression results also confirm the observed positive time trend (p-value = 0.014).

2.6.2.5 Welfare

Figure 2.5 shows the average welfare in the market over time. The average welfare in this treatment fluctuates around 80. There is a weakly significant positive time trend for the NoBarrier treatment, but it is negligible in magnitude.[20] There is no time trend for the Barrier treatment, but values vary between 50 and 65.

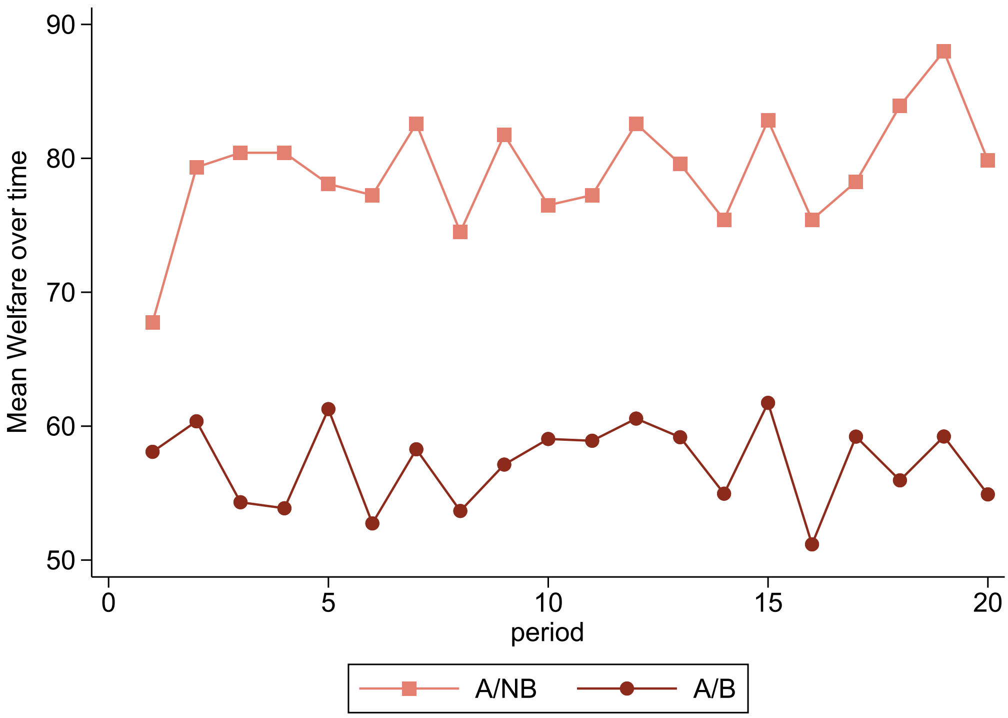


Figure 2.5: Mean welfare in the market over time (Ability).

Result 2.5 *In the Ability scenario, welfare is significantly higher in the NoBarrier treatment compared to the Barrier treatment.*

[20]See Table 2.3 column 5b.

For the Ability scenario, we hypothesize that welfare is lower with the regulatory entry barrier. The related Hypothesis 5 is supported by our data as we see a significant difference in welfare between the treatments in Table 2.3 columns 5a and 5b. Welfare is about 17.84 ECU lower in the Barrier treatment when we control for time effects (p-value < 0.001). This welfare difference is not significantly different from the investment cost for a seller of 16 ECU (p-value $= 0.693$).

Table 2.3: Regression analysis of the ABILITY scenario.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	0.987***	0.558*	-0.185	-0.486	4.116	2.296	-23.61*	-21.68	-21.30***	-17.84***
	(0.114)	(0.245)	(0.361)	(0.538)	(2.578)	(2.785)	(9.687)	(12.05)	(3.223)	(4.648)
period		-0.0437**		-0.0669**		-0.518***		0.880**		0.301*
		(0.0149)		(0.0243)		(0.127)		(0.295)		(0.142)
barrier × period		0.0409**		0.0285		0.172		-0.184		-0.330
		(0.0151)		(0.0361)		(0.148)		(0.592)		(0.229)
Constant	0.900***	1.359***	2.231***	2.934***	12.46***	17.90***	62.28***	53.05***	79.08***	75.92***
	(0.0941)	(0.239)	(0.271)	(0.394)	(2.170)	(2.297)	(6.358)	(9.014)	(1.219)	(2.116)
R^2	.4422791	.5002073	.0022859	.0291385	.0669575	.1699532	.1852371	.2130466	.3441785	.3487981
Observations	960	960	933	933	933	933	1920	1920	960	960

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the BARRIER treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.6.3 Transparency scenario

In this scenario, investment in qualification makes the invested sellers' product quality observable for buyers whereas the quality of products offered by sellers who choose not to invest is not observable. Irrespective of the investment decision in this scenario, sellers can choose any quality level. In the following subsections, we test the pre-registered directed hypotheses 6–10 for this scenario. The theoretical predictions and further possible equilibria for the two treatments in this scenario are explicitly proposed and discussed in more detail in section 2.2 and section 2.A of the Appendix.

2.6.3.1 Investment

Figure 2.6 shows the average number of investments over time in our two experimental treatments. The average investments in the Barrier treatment are close to 2 in all periods. In contrast, in the NoBarrier treatment, the investments show a mixed trend, moving between 1.5 and 0.5 with a slight tendency towards the lower bound. There is no time trend for either treatment.

Result 2.6 *In the Transparency scenario, the average total investment is significantly higher in the Barrier treatment compared to the NoBarrier treatment.*

Table 2.4 presents regressions to investigate the impact of the entry barrier on investment decisions in columns 1a and 1b. We hypothesize that investment levels are the same across treatments. In the experimental data, we find that investments are higher in the Barrier treatment, and Hypothesis 6 is therefore not supported. The existence of an entry barrier significantly increases the number of investments in qualification (p-value < 0.001). The investment level is shifted by about 1 upwards for the Barrier treatment. There is no time trend for this difference.

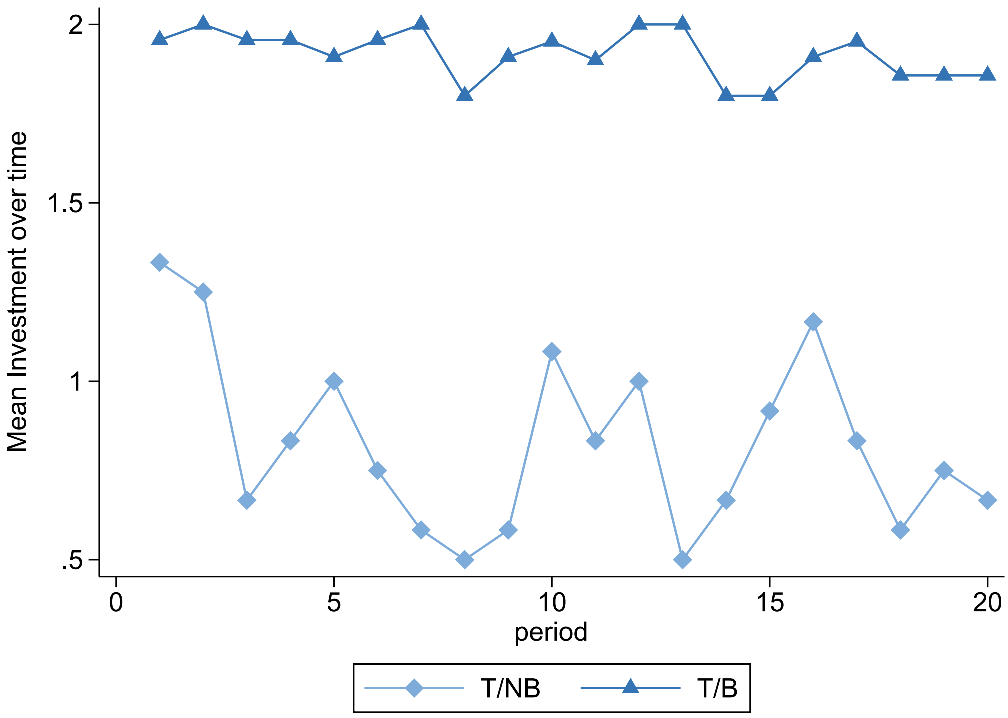


Figure 2.6: Mean number of investments in qualification in the market over time (Transparency).

2.6.3.2 Quality

Figure 2.7 shows the average quality offered in the market over time. We see a significant difference in quality between the treatments, as predicted. Qualities fluctuate around values of 4 for the NoBarrier treatment. For the Barrier treatment, the qualities fluctuate around 8 and 9. For neither treatment do the qualities follow a significant time trend.

Result 2.7 *In the Transparency scenario, the average quality offered in the Barrier treatment is significantly higher than in the NoBarrier treatment.*

The prediction is that the quality will be higher in the Barrier treatment (see Hypothesis 7). Our data shows support for this prediction. We identify a significant difference in the qualities offered between the two treatments (p-value = 0.004). The estimated quality difference between the treatments in the Transparency scenario is significant at 4.52 and 3.56, respectively, when we include a time variable. The

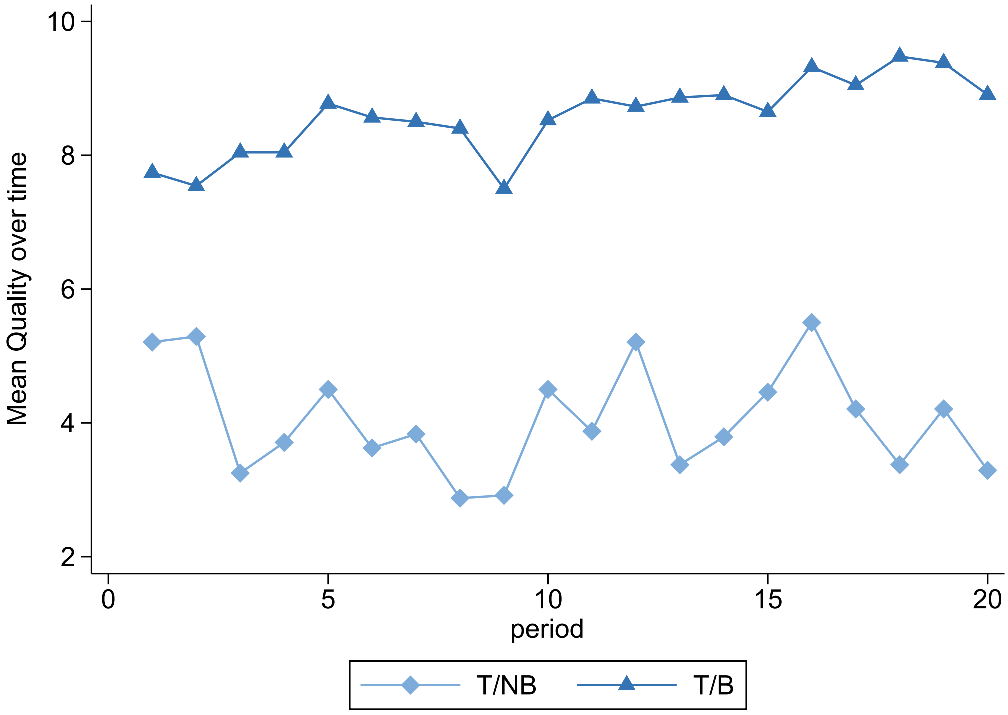


Figure 2.7: Mean quality in the market over time (TRANSPARENCY).

regression results are reported in columns 2a and 2b in Table 2.4. As illustrated by Figure 2.7, sellers choose a higher quality level in the BARRIER treatment.

2.6.3.3 Prices

Figure 2.8 shows the mean price for the products in the market over time. We observe a significant downward trend over all 20 periods for both treatments.[21] In the second half of the experiment, this trend diminishes for the NOBARRIER treatment and disappears for the BARRIER treatment.[22]

Result 2.8 *In the TRANSPARENCY scenario, the average prices in the market are significantly higher in the BARRIER treatment compared to the NOBARRIER treatment.*

We hypothesize that prices in the BARRIER treatment are higher than in the

[21] NOBARRIER: p-value 0.019; BARRIER: p-value< 0.001.

[22] For the NOBARRIER treatment, the estimated time trend is -0.32 ECU per period (p-value $= 0.076$) overall and -0.26 ECU after period 10 (p-value $= 0.017$). For the BARRIER treatment, the overall trend is -0.33 ECU per period (p-value $= 0.008$) and -0.16 ECU after period 10 (p-value $= 0.402$).

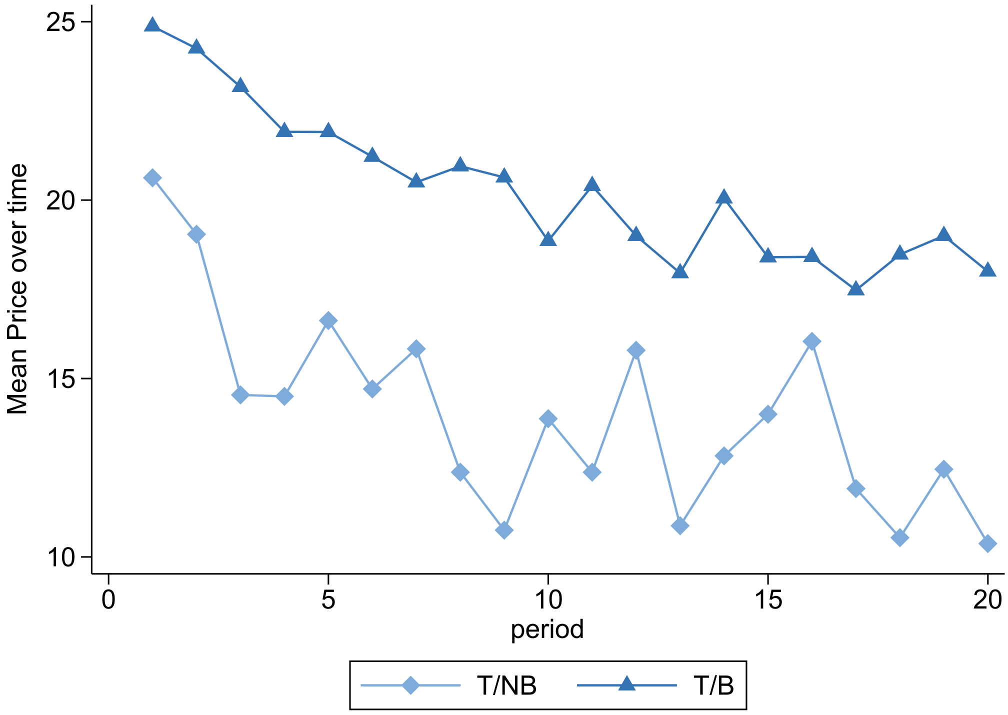


Figure 2.8: Mean price in the market over time (Transparency).

NoBarrier treatment (see Hypothesis 9). Our results presented in columns 3a and 3b in Table 2.4 reflect the prediction for the overall average.[23] When we control for time trends, the difference between the two treatments in this scenario is borderline significant (p-value = 0.104).

2.6.3.4 Consumer surplus

Figure 2.9 presents the average consumer surplus over time. We see a stable and increasing consumer surplus for both treatments, which is in line with the decline in prices documented in section 2.6.3.3.

Result 2.9 *In the Transparency scenario, the average consumer surplus in the market between the Barrier treatment and the NoBarrier treatment do not differ significantly.*

Hypothesis 9 refers to a lower consumer surplus in the NoBarrier treatment.

[23]The average price is 6.34 higher with the barrier (p-value = 0.020).

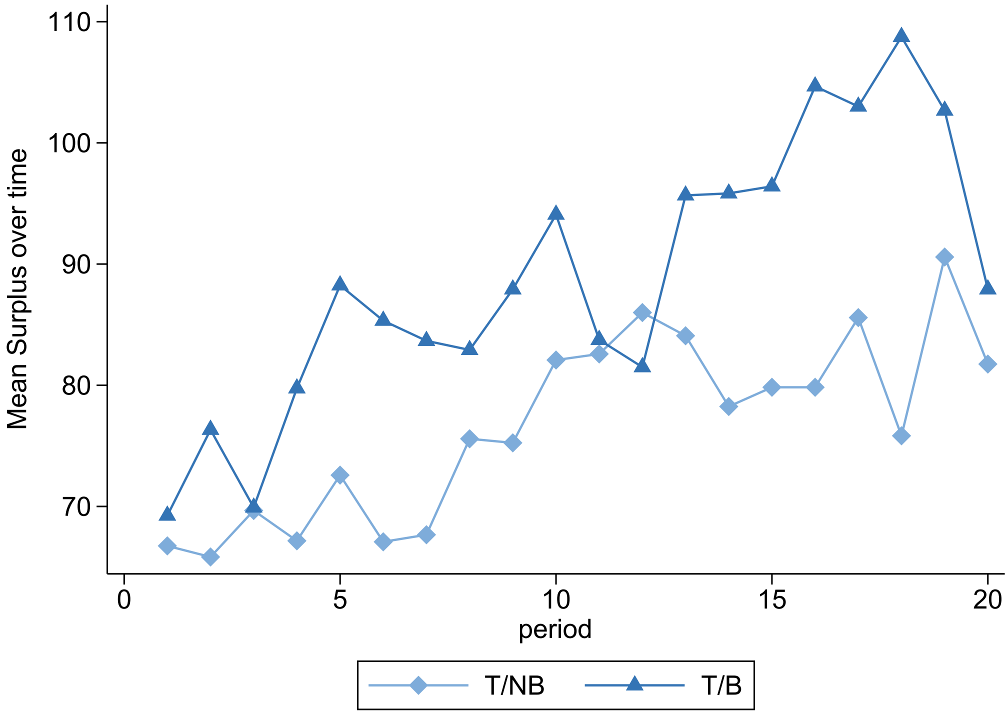


Figure 2.9: Mean consumer surplus in the market over time (Transparency).

The difference between the two treatments is not significant, which is not in line with the hypothesis. Table 2.4 presents the regression results with and without a time variable in columns 4a and 4b.

2.6.3.5 Welfare

Figure 2.10 shows the average welfare in the market over time. In both treatments, welfare is very volatile and the lines are also very similar with some intersections.

Result 2.10 *In the Transparency scenario, the welfare between the Barrier treatment and the NoBarrier treatment is not significantly different.*

Hypothesis 10 formulates welfare to be the same across treatments. This hypothesis is supported by our data. There are no significant differences between the two treatments as reported in columns 5a and 5b in Table 2.4.

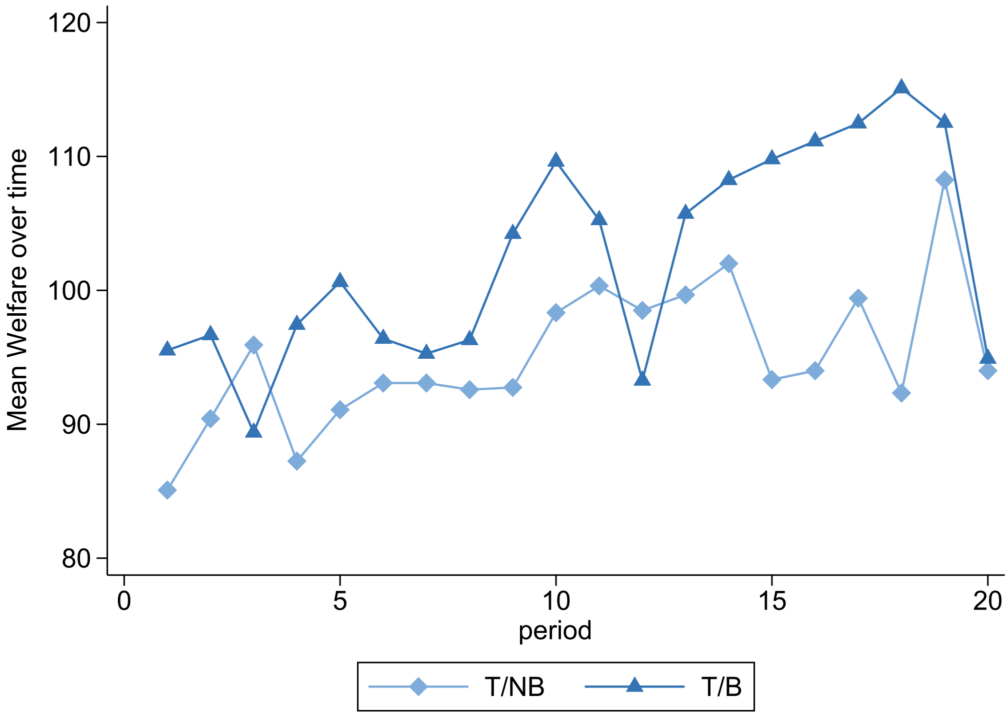


Figure 2.10: Mean welfare in the market over time (Transparency).

Table 2.4: Regression analysis of the TRANSPARENCY scenario.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	0.979***	0.957***	4.520***	3.556**	6.338*	6.354	12.18	6.611	5.175	2.184
	(0.136)	(0.168)	(0.447)	(1.224)	(2.734)	(3.904)	(12.84)	(17.71)	(4.433)	(9.306)
period		-0.0132		-0.0147		-0.322*		1.039		0.548
		(0.0121)		(0.0856)		(0.137)		(0.774)		(0.522)
barrier × period		0.00207		0.0934		-0.00799		0.530		0.285
		(0.0132)		(0.0887)		(0.159)		(1.003)		(0.635)
Constant	0.825***	0.963***	4.050***	4.204***	14.00***	17.38***	76.70***	65.79***	95.07***	89.32***
	(0.0555)	(0.150)	(0.367)	(1.159)	(2.079)	(3.390)	(9.004)	(13.27)	(3.357)	(7.278)
R^2	.4087138	.4171084	.3678263	.3752596	.1259004	.1705459	.0313504	.0811141	.0103244	.0357829
Observations	960	960	913	913	913	913	1920	1920	960	960

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the BARRIER treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.6.4 Summary of hypotheses tests

In the upper part of Table 2.5, we summarize how our results compare to the pre-registered hypotheses. The lower part of the table complements this overview with a summary of our hypothesis tests, reporting the significance of the *barrier* dummy variable.

Overall, we find support for three of five hypotheses in the Ability scenario. Hypothesis 2 that relates to quality differences between the two treatments is not supported and neither is Hypothesis 3 that relates to price differences. For the Transparency scenario, we find support for three of five hypotheses. We find the predicted quality difference (H7) between the treatments as well as the predicted difference in prices (H8). We also find no difference in total welfare across treatments (H10).

Our experimental results support the conclusion that for markets in the Ability scenario, market outcomes are higher when occupational licensing is not enforced as a barrier to entry. For markets in the Transparency scenario, the theoretical inference finds mixed support in our experimental results.

Table 2.5: Summary hypotheses tests.

Hypothesis	Parameter	Supported by data	Type of significant divergence
ABILITY			
H1	Investment	Yes	More investments in BARRIER
H2	Quality	No	No difference
H3	Price	No	No difference
H4	CS	Yes	CS in NOBARRIER higher
H5	Welfare	Yes	Welfare in NOBARRIER higher
TRANSPARENCY			
H6	Investment	No	More investments in BARRIER
H7	Quality	Yes	Quality in BARRIER higher
H8	Price	Yes	Price in BARRIER higher
H9	CS	No	No difference
H10	Welfare	Yes	No difference

	ABILITY			TRANSPARENCY		
	NOBARRIER		BARRIER	NOBARRIER		BARRIER
Investment	0.90	$\overset{***}{<}$	1.89	0.83	$\overset{***}{<}$	1.80
(std. dev.)	(0.24)		(0.16)	(0.14)		(0.31)
Quality	2.23	>	2.03	4.05	$\overset{***}{<}$	8.59
(std. dev.)	(0.70)		(0.64)	(0.93)		(0.60)
Price	12.46	<	16.54	14.00	$\overset{*}{<}$	20.44
(std. dev.)	(5.57)		(3.37)	(5.27)		(4.16)
CS	62.28	$\overset{*}{>}$	38.68	76.70	<	88.88
(std. dev.)	(16.59)		(17.60)	(22.83)		(22.59)
Welfare	79.08	$\overset{***}{>}$	57.78	95.07	<	100.25
(std. dev.)	(3.14)		(7.38)	(8.41)		(7.10)

Notes: Investment: average total number of sellers who invested ($\{0, 1, 2\}$). Quality: average quality offered by a seller (1–10). Price: average price set by a seller (1–60). CS: average consumer surplus in the market. Welfare: average total welfare in the market. We report post-estimation Wald tests, bootstrapped standard errors are clustered at the session level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.7 Analysis of investment subgames

After testing our pre-registered hypotheses in the previous part, we now turn to an explorative analysis of different subgames.[24] Because our games include three decision stages for sellers, followed by the buyer decision, there are several possibilities for subjects to deviate from the equilibrium paths. We now broaden the perspective beyond these equilibrium paths.

By analyzing the subgames, we can evaluate which strategies were played after the investment stage. We are particularly interested in whether the theoretically advantageous strategies of maximum quality differentiation were played, since our hypotheses are mainly based on them. It is also important to analyze and track the paths in detail to identify possible main deviation paths.

Furthermore, we have seen in section 2.6 that the frequency with which the different investment subgames appear in our experimental data is more mixed, or the majority is even shifted toward other subgames than predicted. With the analysis in this section, we also account for this observation.[25]

The structure is as follows: We first consider the different investment subgames and test whether our experimental results for qualities match the predicted qualities in the investment subgames. In a second step, we then turn to the pricing stage.

2.7.1 Qualities

In each investment subgame, we predict quality levels chosen by the two sellers (A and B). These predictions follow from the theoretical considerations of the quality-then-price subgame in Appendix 2.A. In the following two sections, we present the p-values for tests whether the average qualities of our experimental data in the subgames are equal to the predicted qualities. Additionally, we test if the difference between our experimental results for the two sellers matches the difference between

[24] There are four investment subgames in each treatment of the two scenarios. However, at the stage of the quality-then-price subgame, the two subgames with symmetric investment decisions can be considered together because all decisions are made simultaneously after that point.

[25] The frequencies of the investment subgames for each treatment in the two scenarios are included in Tables 2.6 and 2.7 as N *(total)*.

predicted averages.[26]

2.7.1.1 Ability

The test results are presented in Table 2.6. The table contains the predictions, experimental results, and *p*-values for the Wald test of equality of the prediction and the experimental data. The upper part belongs to the NoBarrier treatment. We include the first column $(0,0)$ only for completeness. Due to the experimental design, both qualities are set to 1 in this subgame, and the prediction is also 1 for both for the same reason.

In the second column $(1,0)/(0,1)$, there is also a seller who did not invest and has a default quality of 1. For the seller who invested, the prediction is to serve a higher quality, and we see this in 78 percent of the markets. The higher quality is significantly below the predicted maximum differentiation as a quality difference.

The third column $(1,1)$ shows differentiation in 90 percent of markets in this subgame. The experimental results differ significantly from the predicted values.

The lower part of the table refers to the Barrier treatment. The first column $(0,0)$ reflects the subgame in which no seller invests in qualification. In this case, no seller enters the market and therefore no qualities are selected. This subgame was never reached in the experiment.

The second column $(1,0)/(0,1)$ represents the case where only one seller enters the market and thus becomes a monopolist. In this subgame, we see that the chosen qualities do not differ significantly from the predicted quality of 1. This subgame was reached in 27 markets, representing 11.25 percent of all markets in this treatment.

The subgame in which two sellers invest was reached most often in this treatment (column $(1,1)$). In 51 percent of the markets in this subgame, products were differentiated. However, the amount of quality differentiation in the experiment is

[26] We restrict our analysis to markets where we see quality differentiation and refer to the seller with the higher chosen quality as A in this table. We exclude markets where we see no quality differentiation from the sample tested here. We do this because we are mainly interested in the frequency of differentiation and, when present, to test the magnitude of the differentiation. We report the same tests without restricting to markets with quality differentiation in the Appendix, section 2.B.2.

significantly different from the predictions.

2.7.1.2 Transparency

Table 2.7 presents the test results for this scenario. The table contains the predictions, the experimental results, and the *p*-values for the Wald test of equality between the predictions and the experimental data. The upper part belongs to the NoBarrier treatment.

The first column $(0, 0)$ reflects the subgame in which no seller invests, but both can offer any product quality (where quality is unknown to buyers before the purchase decision). We see quality differentiation in 25 percent of the markets. The lower quality in the case of differentiation is not significantly different from the predicted value of 1. The higher quality, however, is significantly lower than the predicted value, reflecting less than maximum quality differentiation in the experimental market.

When a seller invests in qualification, we see differentiation in 97 percent of the markets. In addition to this very high proportion, the second column $(1, 0)/(0, 1)$ also shows that the differentiation was 6.12, larger than in any other subgame in which we expect differentiation in quality. Testing whether the experimental result is consistent with the prediction, we see that the mean quality for both sellers differs significantly from the predicted values.

For the subgame in which both sellers invest, the prediction is that the sellers offer different levels of quality. In 73 percent of the markets, one seller offers a higher quality than the other. Both qualities are significantly different from the predicted value. The choice of the lower quality seller is relatively high, with a mean quality of 5.73, compared to other subgames in which the prediction is quality differentiation.

Next, consider the lower part of the table, which belongs to the Barrier treatment. The investment subgame in which no seller decides to invest results in no products being offered. There are six markets in which no seller invests in qualification and thus no seller enters the market.

In the second column $(1, 0)/(0, 1)$, one seller enters the market. The quality chosen by this monopolistic seller is not significantly different from the predicted quality of 10. This subgame was reached in 15 percent of the markets in the Barrier treatment of the Transparency scenario.

When both sellers invest in qualification, the prediction for quality choices is again maximum differentiation. In 61 percent of the markets we see different qualities offered to buyers. For both sellers, the experimental result differs significantly from the predicted value. With an average quality of 6.41, the lower quality in this subgame is relatively higher compared to other subgames where sellers are predicted to offer different qualities.

Table 2.6: Point predictions test Ability scenario.

NoBarrier

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	1	1	0	10	1	9	10	1	9
p-value	n/a	n/a	n/a	< 0.001	n/a	< 0.001	< 0.001	< 0.001	< 0.001
Mean	1	1	0	4.39	1	3.39	5.35	2.67	2.67
(Std. Dev.)	-	-	-	*(1.85)*	-	*(1.85)*	*(2.55)*	*(1.58)*	*(1.76)*
N (diff)		n/a			89			46	
N (total)		75			114			51	

Barrier

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	-	-	-	1	-	-	10	1	9
p-value	n/a	n/a	n/a	$= 0.30$	n/a	n/a	< 0.001	< 0.001	< 0.001
Mean	-	-	-	1.63	-	-	4.40	1.52	2.88
(Std. Dev.)	-	-	-	*(1.28)*	-	-	*(1.91)*	*(1.12)*	*(1.81)*
N (diff)		0			27			108	
N (total)		0			27			213	

Notes: We only consider cases in which there is a difference in quality, with the exception of the "lemons" prediction in the $(0,0)$ investment case and with a monopolistic seller. We comprise the two investment combinations in which only one seller invests. In these subgames, we designate the investing seller as seller A in this table. In cases in which both sellers have invested, we designate the one who offers the higher quality as seller A (if applicable). There are no cases when neither of the sellers invested in Barrier. Bootstrapped standard errors are clustered at the session level. We report p-values for post-estimation Wald tests. We test the Null Hypothesis (H_0): mean = prediction. For p-values below 0.05, we reject the null hypothesis, indicating that the observed outcome significantly deviates from the prediction. *Pred* notes the predicted value and *Mean* the experimental result. *N (diff)* reflects the number of markets for which we see the differentiation (or monopolistic sellers) and N *(total)* is the total number of markets in the respective investment subgame.

Table 2.7: Point predictions test TRANSPARENCY scenario.

NOBARRIER

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	*B*	Δ	*A*	*B*	Δ	*A*	*B*	Δ
Pred	10	1	9	10	1	9	10	1	9
p-value	< 0.001	$= 0.13$	< 0.001	< 0.001	$= 0.018$	< 0.001	$= 0.004$	< 0.001	< 0.001
Mean	3.76	1.19	2.57	7.83	1.72	6.12	8.5	5.73	2.77
(Std. Dev.)	*(2.43)*	*(0.51)*	*(2.27)*	*(2.36)*	*(1.61)*	*(2.51)*	*(2.22)*	*(2.59)*	*(1.94)*
N (diff)		21			113			30	
N (total)		83			116			41	

BARRIER

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	*B*	Δ	*A*	*B*	Δ	*A*	*B*	Δ
Pred	-	-	-	10	-	-	10	1	9
p-value	n/a	n/a	n/a	$= 0.23$	n/a	n/a	< 0.001	< 0.001	< 0.001
Mean	-	-	-	9.23	-	-	9.18	6.41	2.77
(Std. Dev.)	-	-	-	*(1.57)*	-	-	*(1.12)*	*(2.04)*	*(2.08)*
N (diff)		0			35			122	
N (total)		6			35			199	

Notes: We only consider cases in which there is a difference in quality, with the exception of the "lemons" prediction in the $(0, 0)$ investment case and with a monopolistic seller. We comprise the two investment combinations in which only one seller invests. In these subgames, we designate the investing seller as seller A in this table. In cases where both sellers have invested, we designate the one who offers the higher quality as seller A (if applicable). Bootstrapped standard errors are clustered at the session level. We report *p*-values for post-estimation Wald tests. We test the Null Hypothesis (H_0): mean = prediction. For *p*-values below 0.05, we reject the null hypothesis, indicating that the observed outcome significantly deviates from the prediction. *Pred* notes the predicted value and *Mean* the experimental result. *N (diff)* reflects the number of markets for which we see the differentiation (or monopolistic sellers) and N *(total)* is the total number of markets in the respective investment subgame.

2.7.2 Prices

Given the numerous quality subgames, we focus in our analysis on price differences as a function of the quality differences in the market. From the theoretical considerations in Appendix 2.A, we know that the difference between the optimal prices for given qualities follows the condition $\Delta p = 2 \cdot \Delta q$.[27] Using this relationship between the difference of the optimal prices given the qualities chosen, we can test whether the price pairs in our experimental data match this relationship to the difference of the qualities in the market. The following two sections present the tests for the different investment subgames. We also include tests for the prices in the other subgames in which there is no differentiation.

2.7.2.1 Ability

In this scenario, quality differentiation is expected in the NoBarrier treatment for the two investment subgames when either one seller invests or when both sellers invest. Figure 2.11 shows the price differences between the two prices in the market, given the quality difference.

The graph on the left refers to the subgame in which one seller invests and can set a quality between 1 and 10 (in this case, the quality of the non-investing seller is 1). The graph shows that the price difference increases mostly according to the predicted slope of 2. We also formally test whether the slope of the fitted values (red line) is equal to 2 and cannot reject this null hypothesis (p-value = 0.574). For high quality differences, we see that the price differences increase more than expected, but overall the prices in these subgames follow the predicted relationship.

The graph for the subgame in which both sellers invest (right) shows a similar pattern. For quality differences above 7, the price difference increases more than for the previous quality difference steps. The slope for the fitted values is also not significantly different from the theoretical slope of 2 (p-value = 0.181). There is no market with maximum differentiation, that is, we do not observe the quality

[27]For a derivation of this relation, see Appendix 2.A.

difference of 9.

When no seller invests, the qualities are 1 for both sellers according to the design, and we predict Bertrand-style price competition. Testing whether the average prices equal the predicted price of 1, we find that the average realized price equals 7.07 and, hence, is significantly (p-value < 0.001, Wald test) above the price of 1.

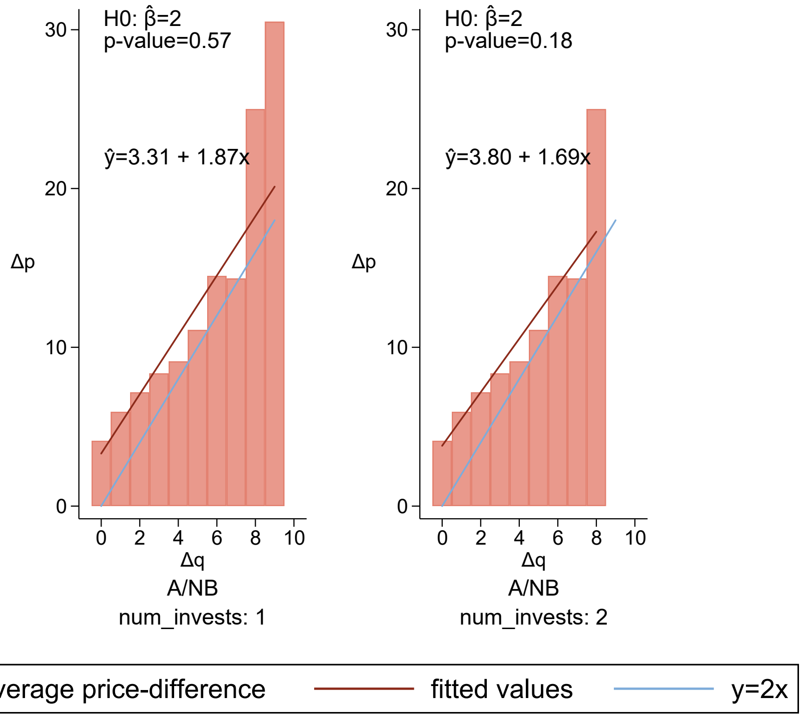


Figure 2.11: Price differences/quality differences in AbilityNoBarrier.

For the Barrier treatment, quality differentiation and an increasing difference in prices is predicted when both sellers invested. The graph in Figure 2.12 shows that the experimental results do not match this prediction. The increase in the price difference of 0.41 is significantly different from the predicted slope of 2.

When one seller has invested, the market is served by a monopolist and the prediction is a price of 21 to achieve the highest profit. We also test whether the average price in the markets is equal to 21. The average price in these markets ($N = 27$) is 18.07 and not significantly below 21 (p-value $= 0.055$).[28] If no seller

[28]When we restrict our attention to those markets in which the quality offered by the monopolist

invested, there is no entry, and, hence products cannot be offered.

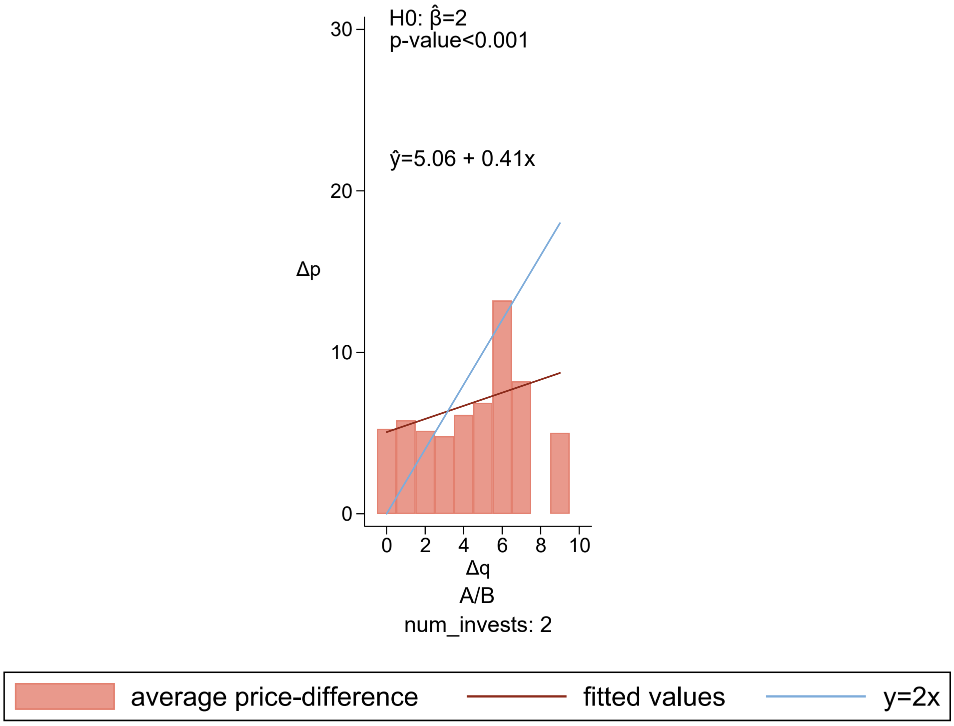


Figure 2.12: Price differences/quality differences in AbilityBarrier.

2.7.2.2 Transparency

For the NoBarrier treatment of this scenario, maximum quality differentiation is the prediction in all three investment subgames. Figure 2.13 shows the price differences given different quality differences for each of them.

When both sellers invested (right graph), the pattern shows an increasing price difference as the quality difference increases, but the slope of the fitted values is significantly less than 2. Particularly noticeable is that the price differences in the high quality differences (7 and 8) are smaller than the predicted values. For smaller quality differences (2 and 3), the price differences are higher than what is predicted.

For the subgame in which one seller invested, the graph in the center shows a similar pattern compared to the subgame in which both sellers invested. Again, the

was 1 ($N = 19$), the average price is 17.95 and significantly (p-value $= 0.017$) below 21.

slope is significantly less than 2, and the price differences are higher than predicted for small quality differences; for large quality differences, the price differences are smaller than predicted.

When no seller has invested (left graph), the slope is steeper than in the other two subgames, but also significantly smaller than 2. For this subgame, we see either small or large quality differences, for which the previous pattern also holds. Whereas the smaller quality differences lead to higher price differences than predicted, the opposite is true for the larger quality differences.

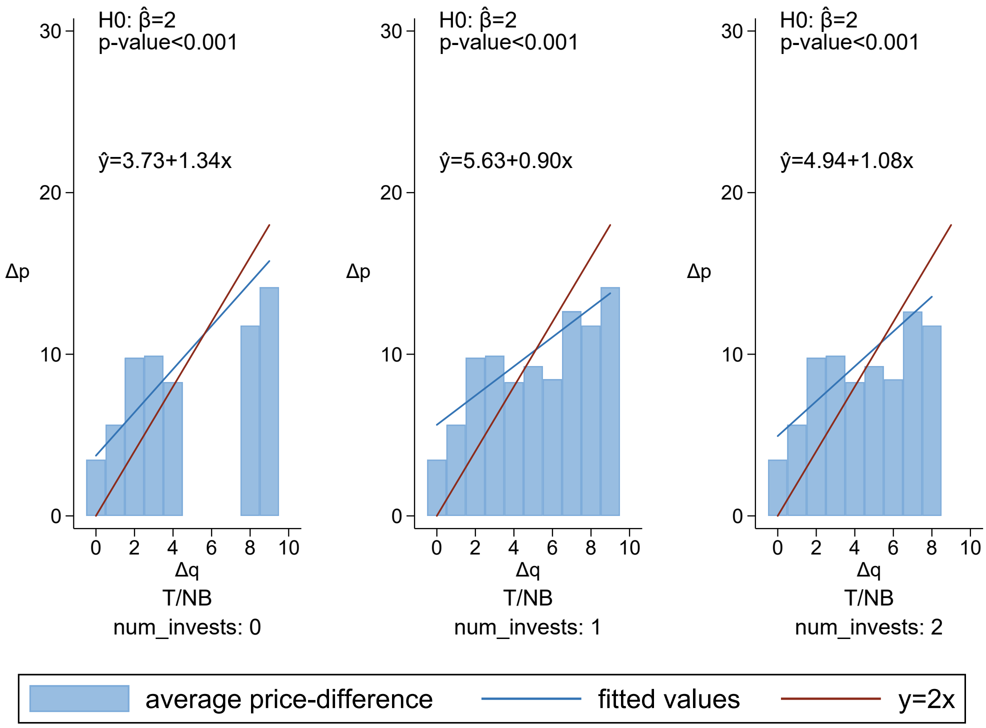


Figure 2.13: Price differences/quality differences in TransparencyNoBarrier.

With the entry barrier, quality differentiation is only predicted when both sellers enter the market. Figure 2.14 shows the graph for the Barrier treatment when both sellers invest. The graph shows the expected pattern of an increasing price difference when the quality difference increases. The slope of the fitted values line is 1.82, which is not much but significantly smaller than 2.

When one seller invests in qualification, the market is served by a monopolist and theory predicts a price of 35 for a quality of 10. The average price in these markets ($N = 35$) is 24.4, which is significantly lower than the predicted value (Wald test, p-value < 0.001). If we restrict the analysis to markets in which the offered quality is 10 ($N = 26$), the average price of 23.69 is also significantly smaller than 35 (Wald test, p-value < 0.001). If no seller invested, there is no entry, and no products can be offered.

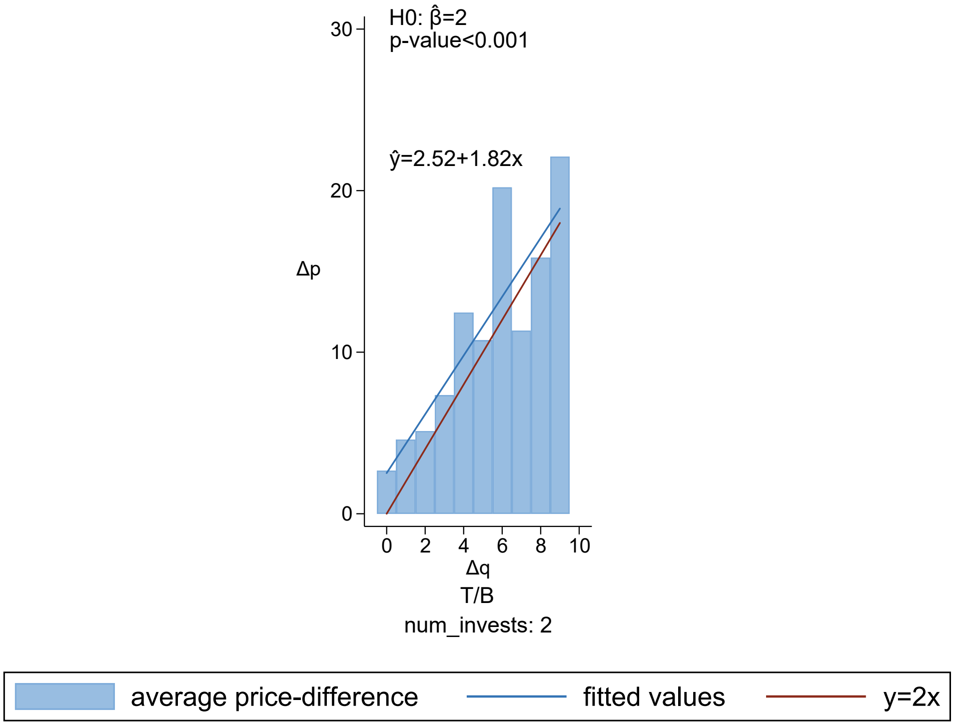


Figure 2.14: Price differences/quality differences in TransparencyBarrier.

2.7.3 Synthesis of investment subgames analysis

When two sellers enter the market, they can either differentiate in quality and serve different groups of buyers, or offer the lowest possible quality and compete for the entire market in Bertrand-style price competition. Theory shows that quality differentiation is mostly advantageous for sellers, especially compared to the zero-profit alternative of Bertrand-style price competition. Maximum quality differentiation is the most common equilibrium in the three quality-then-price subgames. However, this differentiation requires some coordination. In theory, this problem is easy to solve, but in practice (in the experiment) such coordination can be a challenging task for market participants. Therefore, we also look at the share of markets within the subgames in which we can observe quality differentiation.

Our results in section 2.7.1 show that quality differentiation was realized most often in the NoBarrier treatment of the Ability scenario. However, the amount of differentiation was less than half to less than a third of the maximum differentiation. In the Barrier treatment, quality differentiation was observed relatively less often.[29] The amount of differentiation with less than a third of the maximum differentiation was similar to the NoBarrier treatment.

Given these results, there seem to be two (potential) drivers for the differences in the uptake of differentiation between the two treatments. On the one hand, about two-thirds of the observations of differentiation in the NoBarrier treatment come from the subgames with asymmetric investment decisions. This asymmetry – and in particular the experimental design – improves coordination for differentiation because only the seller who invested can choose a higher quality. On the other hand, the mandatory investment in qualification to enter the market in the Barrier treatment may have been much less of a conscious decision for the sellers. Consequently, the sequencing of investment decisions, which may have had a coordination function in the NoBarrier treatment with two sellers who invested,[30] may have been less important in the Barrier treatment.[31] From these observations, we can conclude for the Ability scenario that a conscious decision to invest in qualification, as well as asymmetry in sellers, increases the uptake of differentiation in our experiment.

In the Transparency treatment, not only the share of markets with differentiation is very mixed, but also the magnitude of the realized differentiation for the different subgames. When both sellers invested in qualification, there is differentiation in 73 percent of the markets of the NoBarrier treatment, but the magnitude is

[29]For the NoBarrier treatment, the qualities were differentiated in 82 percent of the markets in this subgame, while this proportion is only 52 percent for the Barrier treatment.

[30]Seller A had the higher quality in 27 of 46 markets in NoBarrier.

[31]Seller A had the higher quality in 49 of 108 markets in Barrier.

less than a third of the maximum differentiation. It is noticeable that the chosen lower quality is considerably higher than in other markets that show this (low) differentiation.[32] The quality choices in this investment subgame show a similar pattern in the Barrier treatment. The level of differentiation is also less than a third of the maximum, and this seems to be driven by a lower quality that is remarkably higher than predicted. At 61 percent, differentiation is realized less often than in the NoBarrier treatment.

In markets where no seller decided to invest in qualification, we observe differentiation in only 25 percent of the markets in the NoBarrier treatment. The level of differentiation is similar to the case where both sellers invested, i.e. less than a third of the maximum differentiation. In contrast, with asymmetric investment decisions, quality differentiation appears in almost all markets (97 percent). In these markets, the level of differentiation is relatively high; it is more than twice as high as in the other two subgames. The average qualities in this subgame are neither significantly shifted to the upper nor to the lower bound of the quality range.

These results lead to the conclusion that, similar to the Ability scenario, asymmetric investment decisions work well as a coordination device in the Transparency scenario.[33] Conversely, the sequential investment decision in the experiment has little, if any, effect on the coordination problem in the Transparency treatment. This conclusion follows from the mixed proportions of differentiation under symmetric investment decisions. We find a higher proportion of markets with differentiation in the subgames where both sellers invested. However, the question remains whether this is (only) related to the visibility of quality (transparency) or whether other motives play a role.[34]

Furthermore, if we compare the two subgames in which both sellers invested, the pattern is not so clear. While in the NoBarrier treatment, seller A chose the

[32]This finding is consistent with the results in Henze et al. (2015) that both sellers tend to choose similarly high qualities, although we find more quality differentiation in our experiment.

[33]In 110 of 113 markets, the higher quality was offered by the seller who invested in qualification.

[34]Because these subgames also involve (sunk) investment costs, more differentiation could also be driven by motives to compensate for a potential loss that would otherwise not occur.

higher quality twice as often as seller B in the symmetric subgames, this ratio is much more balanced in the Barrier treatment. In particular, seller B chose the higher quality even more often in the Barrier treatment.[35] This leads to the same conclusion as in the Ability scenario, which is that it may also matter whether the investment decision was conscious or whether there was more or less no alternative.

Overall, we conclude that quality differentiation mostly works when an (endogenous) asymmetry controls coordination into different seller types. In addition, whether investment is a conscious decision may have an impact on differentiation. Furthermore, the transparency of the quality offered may play a role.

Next, we conclude from section 2.7.2 that pricing is mostly consistent with the equilibrium strategy for the subgames when at least one seller has invested in the NoBarrier treatment of the Ability scenario. However, the price differences are larger than in the equilibrium strategies, especially in the case of high quality differences.

In contrast, price differences are not consistent with the equilibrium strategy and are mostly too small in the Barrier treatment of this scenario. Together with the previous observation that coordination for differentiation seems to be more difficult when investment is less of a conscious decision, these too small price differences in the presence of quality differences lead to the conclusion that differentiation to serve different market segments may have been less of a motive in these markets. Rather, the unobservability of quality for buyers may have led to price competition in this subgame.

In the Transparency scenario, the price differences are not consistent with the equilibrium strategies. In particular, the price differences are too small for the high quality differences in the NoBarrier treatment.

Since the effect of bearing the cost of investing in qualification versus not bearing this cost remains an open question, we can exploit a particular design feature of our experiment. With all due caution when comparing across scenarios in our experiment,

[35]Seller B offered the higher quality in 67 out of 122 markets.

if we take the investment decisions as given, sellers face the same subgame in AbilityNoBarrier when both sellers have invested and in TransparencyNoBarrier when neither has invested. Sellers can choose quality levels from 1 to 10, but the chosen quality level is not observable to buyers. By comparing the results of these two subgames, we can test the effect of (sunk) investment costs compared to the same quality-then-price subgame without investment costs. The only difference at this stage of the game is the sunk cost from the prior investment in qualification in the Ability scenario. Interestingly, the difference between the two quality levels is very similar and not significantly different in these two cases (2.67 and 2.57, p-value = 0.862), but in AbilityNoBarrier both qualities are higher than in TransparencyNoBarrier.[36]

2.8 Discussion

We now discuss and evaluate our findings in a final consideration. From a theoretical perspective, there are positive and negative effects of occupational licensing as an active barrier to entry. This depends not only on the scenario, but also on which market variable should be considered – and from which perspective. Our experimental results emphasize and, crucially, complement the theory by showing where theoretical considerations could be extended.

Our hypothesis tests in section 2.6 lead to the conclusion that our hypotheses are mainly confirmed for the Ability scenario, while it is more ambiguous for the Transparency scenario. The included total welfare is a good starting point for an overall evaluation and comparison of the different treatments.

From this bird's eye view, we can conclude that occupational licensing enforced as an entry barrier leads to worse market outcomes in the Ability scenario. A similarly higher consumer surplus in the NoBarrier treatment supports this conclusion. For the Transparency scenario, the comparison of total welfare gives no clear

[36] For the higher quality the comparison is $5.36 > 3.76$ (Wald test, p-value = 0.084) and for the lower quality $2.67 > 1.19$ (Wald test, p-value = 0.005).

indication, nor does the measured consumer surplus in the markets. At this point, we take a closer look at the different perspectives of sellers and buyers in the market.

The first perspective is that of the seller. As argued earlier, sellers are mostly better off in a differentiation equilibrium than in a Bertrand-style price competition with zero profits. From this perspective, it is important to learn whether differentiation can be achieved when we translate the theory into the experiment. In section 2.7.3, we emphasize that differentiation is easier to achieve with asymmetric investment decisions. From the seller's perspective, occupational licensing enhances differentiation and hence seller profits. However, this is only true if licensing is not a mandatory requirement for entry, but a conscious decision. This means that NoBarrier is preferred in both scenarios from the sellers' point of view. It should also be noted that differentiation makes it possible to address more diverse customer groups, which can be beneficial from a welfare perspective.

The second perspective is the buyer's perspective. For buyers, there are different measures that can be considered. One measure is the average buyer's utility. This measure takes into account the quality and price of the product offered, as well as the different sizes of their effects on the buyer's utility.[37] Using the average buyer[38] keeps this measure neutral to the actual realizations in the experiment. Figure 2.15 shows the comparisons of the average buyer's utilities between the different treatments of the two scenarios. For the Ability treatment, the lighter and darker red bars show a significant (Wald test, p-value = 0.016) difference of the utilities. The average buyer's utility is more than 50 percent higher without the barrier to entry in the Ability scenario. This difference goes in the opposite direction for the Transparency scenario. The average buyer's utility is about 25 percent higher in the Barrier treatment than in the NoBarrier treatment. However, this difference is not significant (Wald test, p-value = 0.117).

A second measure that can be considered is the proportion of buyers served.

[37] Another measure that ignores these different sizes of the effects of the two variables is the price-quality ratio.

[38] The average buyer has a valuation for quality of $\underline{\theta} + (\bar{\theta} - \underline{\theta}) \times \frac{1}{2} = 1 + (4 - 1) \times \frac{1}{2} = 2.5$.

Theoretically, demand is lower in the Barrier treatment for the Transparency scenario, but not for the Ability scenario.[39] This share decreases in our experiment with the entry barrier in both scenarios, but also these shares are high in all treatments (above 94 percent). It is not surprising that we find no significant difference between the share of buyers served in the two treatments in either scenario.[40]

Other measures are consumer surplus and the share of consumer surplus in total welfare. As shown in section 2.6, consumer surplus is higher in the NoBarrier treatment for the Ability scenario, while consumer surplus is not significantly different across treatments in the Transparency scenario. Zooming in on the share of consumer surplus in total welfare, we see that the shares are 78.8 percent for the Ability and 79.8 percent for the Transparency scenario, and thus at similar levels in the NoBarrier treatment.[41] The introduction of the entry barrier reduces the average share of consumer surplus in total welfare to 64.6 percent in the Ability scenario.[42] For the Transparency scenario, the entry barrier increases the share to 87.1 percent.[43]

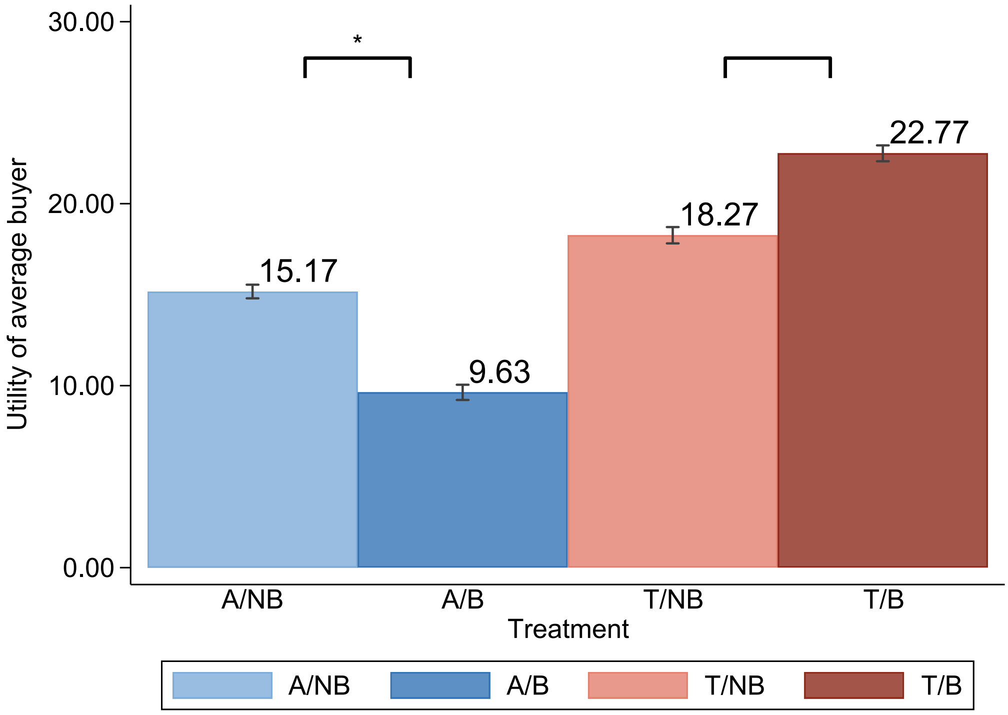
[39] For the theoretical demand served by sellers, see Tables B.1 and B.2 in Appendix 2.A.

[40] For the Ability scenario: p-value = 0.232 (Wald test) and for the Transparency treatment: p-value = 0.380 (Wald test).

[41] We first calculate average values of consumer surplus and welfare per session to address the within-session interdependencies.

[42] The two means are not significantly different (Wald test, p-value = 0.214).

[43] The two means are not significantly different (Wald test, p-value = 0.505).



Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 2.15: Utility of an average buyer.

Taking all perspectives and measures into account, it seems clear that NoBarrier should be preferred for markets in the Ability scenario. Less clear, and depending on the interests of the different market participants, are the reasons for and against occupational licensing as a barrier to entry for markets in the Transparency scenario.

2.9 Conclusion

In this laboratory market experiment, we study if and how occupational licensing enforced as market entry barrier affects market outcomes. Given different consequences that can result from obtaining the occupational license, we differentiate between the two scenarios Ability and Transparency. Using our experimental data, we test the theory-based hypotheses that formulate statements about the realized values of different market variables for our two treatments (NoBarrier and Barrier). From these hypothesis tests, we can conclude that a barrier to entry harms overall welfare in the Ability scenario markets, whereas the results are ambiguous for the Transparency scenario. We further examine the different subgames to see how consistent the realized values are with the predicted strategy paths. Using this analysis, we see that the option of occupational licensing can enhance sellers' coordination for vertical differentiation, and accordingly, not having an entry barrier associated with the license would be preferred from their perspective. From the buyers' perspective, preferences depend more on the scenario. Whereas we find that buyers are clearly better off without an entry barrier in the Ability scenario, we can only identify a tendency toward a preference for an entry barrier in the Transparency scenario.

Translating these findings into policy implications, our main point is that the scenario matters. When policymakers consider introducing or reinstating occupational licensing as a barrier to entry, they should consider the effects of obtaining the license. That is, a classification of markets between the two scenarios, interpreted as bounds of a continuous range, can indicate whether the entry barrier could have positive effects for consumers while being welfare neutral, or whether it is actually harmful from multiple perspectives. Accordingly, for markets that can be classified as similar to markets in the Ability scenario, market entry barriers related to occupational licensing should be repealed.

We find that deliberately choosing an occupational license can be good for differentiation because it can help with coordination. It is interesting to explore this idea further. It might be worth considering offering more than one professional license. This would combine the idea of a minimum standard by limiting market access to the existence of a license with the option of consciously acquiring another license, which in turn would have a positive effect on coordination. However, an evaluation of these ideas goes beyond what we can explain with our experimental results and remains an open question for further research.

Appendices

Appendices

2.A Theoretical predictions

We discuss the derivation of the equilibria in the subgames for the market outlined in section 2.2.1. We proceed by backward induction.

Note that some of the potential scenarios that follow the sequential investment decision by the two sellers (that is, the quality-then-price subgame) are well known from the literature. The case in which both sellers choose qualities that are not observed by the buyers has been investigated by Akerlof (1970), whereas Shaked and Sutton (1982) have analyzed the case in which both sellers choose qualities that can be observed by buyers. For our concrete parameters, the equilibrium outcomes for these cases (among others) have been derived by Henze et al. (2015).[44] The following builds on their analysis where applicable.

We start with the Ability scenario and then proceed with the equilibria for the Transparency scenario.

2.A.1 Ability scenario

In the Ability scenario, investments enable sellers to choose a (higher-than-minimum) quality, which is, however, not observed by buyers. For sellers that do not invest, the product can only be of the lowest quality, which is known to buyers. In the quality-then-price subgame, the investment costs for qualification are sunk. We distinguish between the cases for the two treatments NoBarrier and Barrier.

2.A.1.1 Quality-then-price subgame

Given the possible investment decisions, there are three (qualitatively different) scenarios: (i) both sellers invested, (ii) only one seller invested, and (iii) no seller invested. The informational consequences for buyers may differ in these cases, depending on the treatment. In what follows, we derive perfect Bayesian equilibria in

[44] See their (online) Appendix *A*.

the different scenarios. In a perfect Bayesian equilibrium, all players' strategies must be optimal given beliefs, and beliefs must be derived from equilibrium strategies using Bayes' rule whenever possible.

No entry barrier

If sellers do not have to invest to enter the market, the profits in the three cases depending on investments can be derived as follows:

Case (i) (both sellers invested, both qualities are not observable for buyers): First, the so-called "lemons" outcome is an equilibrium. To see this, consider the following profile of strategies and beliefs. Both sellers choose the lowest quality (that is, $q_i = q_j = 1$, $i, j \in \{A, B\}$, $i \neq j$), and they set prices equal to marginal costs (that is, $p_i = p_j = 1$). Buyers believe that the quality that is offered by both sellers is at the lowest possible level. With regard to buyers' out-of-equilibrium beliefs, assume that whenever buyers observe a price that is different from marginal costs, they believe that the product is of lowest quality. Given these beliefs, no seller has an incentive to unilaterally deviate, and beliefs are derived from equilibrium strategies. We can thus conclude that this profile of strategies and beliefs constitutes a perfect Bayesian equilibrium.

Beside this inefficient equilibrium outcome, it can be shown that equilibria exist in which sellers signal their diverging qualities by setting different prices.[45] More precisely, we argue that the outcomes under full information (see section 2.A.2 below) can be established as equilibrium qualities and prices, that is, $q_A^* = 10$, $q_B^* = 1$, $p_A^* = 28$, and $p_B^* = 10$ (and vice versa). We also comment on other possible equilibrium results and why they may be less plausible.

Consider the following profile of strategies and beliefs, where we distinguish between scenarios on and off the equilibrium path ($i, j \in \{A, B\}$, $i \neq j$):

- Seller i's strategy:
    - On the equilibrium path: The seller chooses quality $q_i = 10$ and sets price

[45]The following analysis closely follows Appendix *A.2* in Henze et al. (2015).

$p_i(10,1) = 28$.

- Off the equilibrium path: For qualities $q_i = 10$ and $q_j > 1$, seller i sets price $p_i(10, q_j) = 28$, independent of seller j's product quality. For any quality $q_i < 10$, the seller sets price $p_i(q_i, q_j) = (19 + q_i)/2$, independent of seller j's product quality.

- Seller j's strategy:
  - On the equilibrium path: The seller chooses quality $q_j = 1$ and sets price $p_j(1, 10) = 10$.
  - Off the equilibrium path: For quality $q_i = 10$, seller j sets price $p_j(10, q_j) = (19 + q_j)/2$ for any $q_j > 1$; for quality $q_i < 10$, seller j sets price $p_j = 28$, independent of its own product quality.

- Buyers' strategies:
  - On the equilibrium path: For prices $p_i = 28$ and $p_j = 10$, buyers with valuation $\theta \geq 2$ buy from seller i, whereas the rest buy from seller j.
  - Off the equilibrium path: For prices $p_i = 28$ and $p_j = p \notin \{10, 28\}$, buyers with valuation $\theta \geq (28 - p_j)/9$ buy from seller i, whereas the rest buy from seller j. For any other price pair, buyers buy from the cheaper seller as long as $\min\{p_i, p_j\} \leq 20 + \theta$ (or randomize if prices are equal); they do not buy otherwise.

- Buyer beliefs with regard to qualities $\mathfrak{q}(p_i, p_j) = (\mathfrak{q}_i, \mathfrak{q}_j)$:
  - On the equilibrium path: For prices $p_i = 28$ and $p_j = 10$, buyers hold beliefs $\mathfrak{q}(28, 10) = (10, 1)$.
  - Off the equilibrium path: For prices $p_i = 28$ and $p_j = p \notin \{10, 28\}$, buyers hold beliefs $\mathfrak{q}(28, p) = (10, 1)$. For prices $p_i = p_j = 28$, or for $p_i \neq 28$ and $p_j \neq 28$, buyers hold beliefs $\mathfrak{q}(p_i, p_j) = (1, 1)$.

We now check whether this profile constitutes a perfect Bayesian equilibrium. Sellers' profits equal $\pi_i^* = 48$ and $\pi_j^* = 12$ given the above strategy and belief profile. In the following analysis, we consider deviations on and off the equilibrium path together where possible.

Consider deviations with regard to prices first. Suppose that any pair of qualities (q_i, q_j) was chosen in the previous stage. There are two qualitatively different cases that must be considered: $q_i = 10$ and $q_i \neq 10$. For $(10, q)$, seller i must set $p_i = 28$ according to its equilibrium strategy. As long as seller j sets a price $p_j \neq 28$, buyers beliefs are independent of the price and equal $\mathbf{q} = (10, 1)$. Generally speaking, the buyer who is indifferent between the two sellers and has valuation for quality of $\tilde{\theta}$ is given as

$$20 + \tilde{\theta}\mathbf{q}_i - p_i = 20 + \tilde{\theta}\mathbf{q}_j - p_j \Leftrightarrow \tilde{\theta} = \frac{p_i - p_j}{\mathbf{q}_i - \mathbf{q}_j}.$$

Assume for the moment that $1 \leq \tilde{\theta} \leq 4$ and buyers believe that $\mathbf{q}_i > \mathbf{q}_j$. Then, the probability that a given buyer prefers the product from seller i amounts to $1 - (\tilde{\theta} - 1)/3$; the respective probability that seller j is preferred is given by $(\tilde{\theta} - 1)/3$. The demands for the two sellers amount to

$$D_i\left(p_i, \mathbf{q}_i; p_j, \mathbf{q}_j\right) = \frac{4}{3}\left(4 - \frac{p_i - p_j}{\mathbf{q}_i - \mathbf{q}_j}\right)$$

and

$$D_j\left(p_j, \mathbf{q}_j; p_i, \mathbf{q}_i\right) = \frac{4}{3}\left(\frac{p_i - p_j}{\mathbf{q}_i - \mathbf{q}_j} - 1\right).$$

From the maximization problem of seller i

$$\max_{p_i} \pi_i = (p_i - q_i)\, D_i\left(p_i, \mathbf{q}_i; p_j, \mathbf{q}_j\right),$$

we can derive the best-response functions

$$p_i(p_j) = \frac{p_j + q_i + 4\left(\mathbf{q}_i - \mathbf{q}_j\right)}{2} \tag{B.1}$$

and

$$p_j\,(p_i) = \frac{p_i + q_j - \boldsymbol{q}_i + \boldsymbol{q}_j}{2}. \tag{B.2}$$

In the case we are interested in, this means that seller j sets an optimal price $p_j = (19+q)/2$, as specified in its strategy profile. The resulting profit amounts to $\pi_j = 4((19+q)/2-q)((28-(19+q)/2)/9-1)/3 > 0$. By contrast, if seller j deviates and sets a price $p_j = 28$, beliefs change and equal $\boldsymbol{q} = (1,1)$. Because even a buyer with the highest valuation for quality would not buy at this price $(20 + 4 \cdot 1 < 28)$, the deviation profit is zero.

Turning to seller i, the best response given by Equation (B.2) for seller j in the concrete case results in a profit of $\pi_i = 4(28-10)(4-(28-(19+q)/2)/9) \geq 48$ (where $\partial\pi_i/\partial q > 0$). Now if seller i sets a deviating price $p_i \neq 28$, buyers hold beliefs $\boldsymbol{q} = (1,1)$ for any q and $p_j(10,q) \neq 28$. Given the lack of quality differentiation from the buyers' point of view, there is Bertrand-style competition, and seller i optimally slightly undercuts seller j and sets $p_i = (19+q)/2 - \epsilon$. Because $\partial p_i/\partial q > 0$, the most favorable outcome is $q = 10$ and, hence, $p_i = 29/2 - \epsilon$. In this case, seller i serves the whole market and, for $\epsilon \to 0$, makes a profit of $4(29/2-10) = 18 < 48$.

We turn to the case where $q_i = q' < 10$ and $q_j = q$. According to its equilibrium strategy, seller j sets price $p_j = 28$. For any price $p_i \neq 28$, buyers hold beliefs $\boldsymbol{q} = (1,10)$; from equation (B.2), seller i sets its best-response price $p_i = (19+q')/2$ as specified in the strategy profile. The resulting profit is $\pi_i = 4((19+q')/2 - q')((28-(19+q')/2)/9-1)/3 > 0$. On the other hand, if the seller i deviates and sets a price $p_i = 28$, the buyers have beliefs $\boldsymbol{q} = (1,1)$. Again, since not even the buyer with the highest valuation of quality would buy at this price $(20 + 4 \cdot 1 < 28)$, the deviation profit is zero.

Next consider seller j. Sticking to the price $p_j = 28$ from the strategy profile gives a profit of $\pi_j = 4(28-q)(4-(28-(19+q')/2)/9)/3$. Because for any q', $p_i(q',q) \neq 28$ and hence $\boldsymbol{q} = (1,1)$ for any price $p_j \neq 28$. The best deviation of seller j is to slightly undercut seller i, so the price is $p_j = (19+q')/2 - \epsilon$, resulting in a deviation profit $\pi_j'' = 4((19+q')/2 - q)$ as $\epsilon \to 0$. Note that the difference

$\pi_j(q', q) - \pi_j''(q', q) = (934 - 2q'(q-1) + 38q)/27$ is decreasing in q' and increasing in q. Since for the lower bound it holds that $\pi_j(9, 1) - \pi_j''(9, 1) = 36 > 0$, deviation is never profitable.

Next consider the case in which a seller unilaterally deviates with regard to its quality and sets $q \neq q_i^*$. As a consequence, in the price-setting stage, sellers set prices $p_i = (19 + q_i)/2$ and $p_j = 28$ and buyers hold beliefs $\mathfrak{q}(p_i, p_j) = (1, 10)$. Seller i thus makes a deviation profit of $\pi_i = 4((28-(19+q)/2)/9-1)((19+q)/2-q) = (19-q)^2/27$. Because $\partial\pi_i/\partial q < 0$, seller j cannot profitably deviate from $q_j^* = 1$. For seller i, setting $q = 1$ yields a deviation profit of $12 < 48$. Hence, deviation is not profitable for either seller.

Note that the above profile is not the only equilibrium that features quality and price differentiation. For example, an equilibrium with qualities $q_i^* = 10$ and $q_j^* = 1$, prices $p_i^* = 32$ and $p_j^* = 12$, and respective beliefs as before exists. In this case, both sellers make higher profits, that is, $\pi_i^* \approx 52$ and $\pi_j^* \approx 18$.

As a consequence, even the profit of the low-quality seller can exceed the fixed cost of entry (market barrier), which means that entry can be profitable. However, such equilibria can be eliminated using the intuitive criterion (Cho and Kreps, 1987). In the example, buyers believe that the seller i is offering low quality when they observe a price $p_i \neq 32$. Consider a deviation price of 29. The low-quality seller could never benefit from this deviation, so buyers should infer that the deviation came from the high-quality seller. When buyers' beliefs are altered in this way, it follows that the high-quality seller benefits by deviating to the price of $29. Infact, theresultingprofitisabout 53.5$, which is more than in the equilibrium candidate. As a result, the above equilibrium (and all similar equilibria with prices higher than in the full information case) can be ruled out.

Case (ii) (only one seller invested, its quality is not observable by buyers, the other sellers' quality is the lowest possible and observable by buyers): Now consider again the possibility of differentiated levels of quality. This scenario is a reduced version of the previous case. Buyers know that the seller that has not invested (seller B) offers low quality, that is, their belief is $q_B = 1$. We again show that qualities $q_A^* = 10$ and $q_B^* = 1$ and prices $p_A^* = 28$ and $p_B^* = 10$ can be established as part of a perfect Bayesian equilibrium.

Consider the following profile of strategies and beliefs, where we distinguish between scenarios on and off the equilibrium path:

- Seller A's strategy:
  - On the equilibrium path: The seller chooses quality $q_A = 10$ and sets price $p_A(10, 1) = 28$.
  - Off the equilibrium path: For any quality $q_A < 10$, the seller sets price $p_A = q_B = 1$.
- Seller B's strategy:
  - On the equilibrium path: The seller chooses quality $q_B = 1$ and sets price $p_B(1, 10) = 10$.
  - Off the equilibrium path: For quality $q_A < 10$, seller B sets price $p_B = q_B = 1$.
- Buyers' strategies:
  - On the equilibrium path: For prices $p_A = 28$ and $p_B = 10$, buyers with valuation $\theta \geq 2$ buy from seller A, whereas the rest buy from seller B.
  - Off the equilibrium path: For prices $p_A = 28$ and $p_B = p \notin \{10\}$, buyers with valuation $\theta \geq (28 - p_B)/9$ buy from seller A, whereas the rest buy from seller B. For any other price pair, buyers buy from the cheaper seller as long as $\min\{p_A, p_B\} \leq 20 + \theta$ (or randomize if prices are equal); they do not buy otherwise.
- Buyer beliefs with regard to quality q_B, $q_A(p_A, p_B)$:
  - On the equilibrium path: For prices $p_A = 28$ and $p_B = 10$, buyers hold belief $q_A(28, 10) = 10$ and know that $q_B = 1$.

– Off the equilibrium path: For prices $p_A = 28$ and $p_B = p \notin \{10\}$, buyers hold belief $q_A(28, p) = 10$. For any other price pair, buyers hold belief $q_A(p_A, p_B) = 1$.

The incentives to stick to this profile are very similar to those in the preceding case and are therefore omitted. The refinement argument with regard to other equilibria also holds here.

Case (iii) (no seller invested, both sellers' quality is the lowest possible and observable by buyers): In this case, both sellers offer the same low quality $q_i = q_j = 1$ $(i, j \in \{A, B\},\ i \neq j)$. Due to the lack of differentiation, there is Bertrand-style competition, which implies that sellers set prices $p_i = p_j = 1$ and make zero profits.

Entry barrier

If the investment is a prerequisite to enter the market in the first place, the outcomes in the three scenarios can be characterized as follows:

Case (i) (both sellers invested, both qualities are not observable for buyers): Because buyers cannot observe the quality chosen by the sellers, the possible equilibrium outcomes are the same as in the case with no barrier.

Case (ii) (only one seller invested, its quality is not observable by buyers): The single seller that invested becomes a monopolist (seller A). The monopolist cannot credibly commit to a certain (high) quality level. Hence, for any price chosen, the monopolist has an incentive to offer the lowest quality possible to save on costs. In this case, the highest price that can be set to ensure all buyers purchase equals $p = 21$. To check whether a higher price (and, hence, less demand) is profitable, consider the monopolist's profit of

$$\pi = (p-1)\frac{4}{3}(24-p)\,.$$

Note that $\partial\pi/\partial p = 4(25-2p)/3 < 0$ because $p \geq 21$ must hold. As a result, the monopolist serves the whole market at a price $p = 21$ and makes a profit of $\pi = 80$.

Case (iii) (no seller invested, both sellers' quality is the lowest possible and observable by buyers): Both sellers are not active in the market and therefore make zero profit.

2.A.1.2 Investment decisions

We now consider the previously identified subgames as given to determine the investment decisions of the sellers. While sellers choose simultaneously in the quality choice and price setting stages, the investment decision is sequential.

No entry barrier

Note first that it is not optimal for any seller to invest if there is the "lemons" outcome in the quality-then-price subgame.

If there is differentiation as under full information in the subsequent subgame (that is, maximum quality differentiation), it is optimal for exactly one of the two sellers to invest. If none of the sellers invest, they make zero profits due to Bertrand-style competition. If both sellers invest and there is maximum differentiation in quality levels, the low quality can be offered without investment. It is better for the low-quality seller not to invest in the first place to save costs.

With regard to investment incentives, note that seller A has a first-mover advantage. If it invests, it becomes the high-quality seller and makes higher profits than its competitor in the quality-then-price subgame. When we account for the investment costs of 16, the higher profits in the quality-then-price subgame justify the initial investment in Ability. The predicted equilibrium strategies for sellers along the equilibrium path are as follows:

- $invest_A = 0$, $q_A = 1$, $p_A(p_B) = 1$ and $invest_B = 0$, $q_B = 1$, $p_B(p_A) = 1$ ("lemons") if buyer beliefs are such that sellers always offer the lowest possible quality

  or

- $invest_A = 1$, $q_A = 10$, $p_A(p_B) = 28$ and $invest_B = 0$, $q_B = 1$, $p_B(p_A) = 10$ (differentiation) if buyer beliefs are as formulated in case (ii) above.

Entry barrier

In this case, sellers that do not invest make zero profits. In the "lemons" outcome and under maximum quality differentiation in the quality-then-price subgame, it is not optimal for at least one seller to enter the market in the first place. As a consequence, only one seller enters to become the monopolist. This is seller A given the sequential decision making. The predicted equilibrium strategies for sellers along the equilibrium path are as follows:

- $invest_A = 1$, $q_A = 1$, $p_A = 21$ and $invest_B = 0$, $q_B = -$, $p_B = -$ (monopoly).

Comparison

As pointed out, the "lemons" outcome, which does not require investments, always constitutes an equilibrium if both sellers are active in the market. The lowest quality, however, is also the optimal choice by a monopolist in the case with an entry barrier, which means that total welfare is lower due to the investment costs. Under maximum quality differentiation, average quality is higher and average price is lower, resulting in higher consumer surplus and total welfare.

In Table B.1, we summarize the predicted strategies. We include the demand for both sellers' products.

Table B.1: Equilibria Ability scenario.

	NoBarrier				Barrier	
Equilibrium	Lemons Market		Differentiation		Monopoly	
Seller	A	B	A	B	A	B
Investment	0	0	1	0	1	0
Quality	1	1	10	1	1	-
Price	1	1	28	10	21	-
Demand	2	2	$\frac{8}{3}$	$\frac{4}{3}$	4	-
Profit	0	0	32	12	64	-

2.A.2 Transparency scenario

In the Transparency scenario, investment enables the buyers to observe the quality which is set by the seller. For sellers that did not invest, the offered quality remains to be not observable by buyers. For the quality-then-price subgame, we ignore the potential costs of investment in qualification that result from entering the different investment cases. We acknowledge this component in a last step. We now distinguish between the cases for the two treatments NoBarrier and Barrier.

2.A.2.1 Quality-then-price subgame

Again, given the possible investment decisions, there are three (qualitatively different) scenarios: (i) both sellers invested, (ii) only one seller invested, and (iii) no seller invested. The informational consequences for buyers may differ in these cases, depending on the treatment. For case (i), we derive the subgame-perfect Nash equilibrium. For cases (ii) and (iii), we derive perfect Bayesian equilibria.

No entry barrier

If sellers do not have to invest to enter the market, the respective profits in the three cases can be derived as follows.

Case (i) (both sellers invested, both qualities are observable for buyers): In the price-setting stage, assume without loss of generality that quality levels were chosen, such that $q_i > q_j$ ($i, j \in \{A, B\}$, $i \neq j$).[46] The buyer who is indifferent between the two sellers and has a valuation for quality of $\tilde{\theta}$ is given as

$$20 + \tilde{\theta}q_i - p_i = 20 + \tilde{\theta}q_j - p_j \Leftrightarrow \tilde{\theta} = \frac{p_i - p_j}{q_i - q_j}.$$

Assume for the moment that $1 \leq \tilde{\theta} \leq 4$. Then, the probability that a given buyer prefers the product from seller i amounts to $1 - (\tilde{\theta} - 1)/3$; the respective probability that seller j is preferred is given by $(\tilde{\theta} - 1)/3$. The demands for the two sellers then

[46]Note that when $q_i = q_j$, all buyers buy from the cheaper seller (see below). If prices are the same, it is assumed that that buyers randomize between the two sellers.

amount to

$$D_i\left(p_i, q_i; p_j, q_j\right) = \frac{4}{3}\left(4 - \frac{p_i - p_j}{q_i - q_j}\right)$$

and

$$D_j\left(p_j, q_j; p_i, q_i\right) = \frac{4}{3}\left(\frac{p_i - p_j}{q_i - q_j} - 1\right).$$

From the maximization problem of seller i

$$\max_{p_i} \pi_i = \left(p_i - q_i\right) D_i\left(p_i, q_i; p_j, q_j\right),$$

we can derive the best response functions

$$p_i\left(p_j\right) = \frac{p_j + q_i + 4\left(q_i - q_j\right)}{2} \qquad \text{and} \qquad p_j\left(p_i\right) = \frac{p_i + q_j - q_i + q_j}{2}.$$

As a consequence, the equilibrium prices can be derived as

$$p_i = 3q_i - 2q_j \qquad \text{and} \qquad p_j = q_i.$$ [47]

In the quality-setting stage, sellers anticipate to make profits of

$$\pi_i\left(q_i; q_j\right) = \frac{16(q_i - q_j)}{3} \qquad \text{and} \qquad \pi_j\left(q_j; q_i\right)) = \frac{4(q_i - q_j)}{3}.$$

Note that $\partial\pi_i/\partial q_i > 0$ and $\partial\pi_j/\partial q_j < 0$. Hence, seller i sets the highest possible quality $q_i = 10$, whereas seller j sets the minimum quality $q_j = 1$. The resulting prices and profits from setting qualities and prices are then given as $p_i = 28$, $p_j = 10$, $\pi_i = 48$, and $\pi_j = 12$.[48]

Case (ii) (only one seller invested, its quality is observable by buyers, the other sellers' quality is not observable by buyers): Note first that the inefficient "lemons"

[47] Given our parameters for $\bar{\theta}$ and $\underline{\theta}$ and the best-response functions $p_i^* = 3q_i - 2q_j$ and $p_j^* = q_i$, we can derive $\Delta p = p_i^* - p_j^* = 2 \cdot (q_i - q_j) = 2 \cdot \Delta q$ to describe the relationship of equilibrium prices and qualities in the markets. We use this relation in section 2.7.2.

[48] Furthermore, an equilibrium in mixed strategies exists, where sellers mix between the two quality levels and make profits of 8.8.

outcome is not an equilibrium here. Because sellers make zero profits in this equilibrium, the seller that invested would not have done so in the first place. Again, an equilibrium as under full information can be established as equilibrium qualities and prices, where the invested seller (seller A) secures the higher profit by offering the high quality: $q_A^* = 10$, $q_B^* = 1$, $p_A^* = 28$, and $p_B^* = 10$.

Consider the following profile of strategies and beliefs, where we distinguish between scenarios on and off the equilibrium path:

- Seller A's strategy:
  - On the equilibrium path: The seller chooses quality $q_A = 10$ and sets price $p_A(10, 1) = 28$.
  - Off the equilibrium path: For qualities $q_A = 10$ and $q_B > 1$, seller A sets price $p_A(10, q_B) = 28$, independent of seller B's product quality. For any quality $q_A < 10$, the seller sets price $p_A(q_A, q_B) = (19+q_A)/2$, independent of seller B's product quality.
- Seller B's strategy:
  - On the equilibrium path: The seller chooses quality $q_B = 1$ and sets price $p_B(1, 10) = 10$.
  - Off the equilibrium path: For quality $q_A = 10$, seller B sets price $p_B(10, q_B) = (19 + q_B)/2$ for any $q_B > 1$; for quality $q_A < 10$, seller B sets price $p_A = 28$, independent of its own product quality.
- Buyers' strategies:
  - On the equilibrium path: For prices $p_A = 28$ and $p_B = 10$, buyers with valuation $\theta \geq 2$ buy from seller A, whereas the rest buy from seller B.
  - Off the equilibrium path: For prices $p_A = 28$ and $p_B = p \notin \{10\}$, buyers with valuation $\theta \geq (28 - p_B)/9$ buy from seller A, whereas the rest buy from seller B. For any other price pair, buyers buy from the cheaper seller

as long as $\min\{p_A, p_B\} \leq 20 + \theta$ (or randomize if prices are equal); they do not buy otherwise.

- Buyer beliefs with regard to qualities $\mathit{q}(p_A, p_B) = (q_A, \mathit{q}_B)$:
  - On the equilibrium path: For prices $p_A = 28$ and $p_B = 10$, buyers hold belief $\mathit{q}_B(28, 10) = 1$ and know that $q_A = 10$.
  - Off the equilibrium path: For prices $p_A = 28$ and $p_B = p \notin \{10\}$, buyers hold beliefs $\mathit{q}(28, p) = (q_A, 1)$. For prices $p_A = p' \neq 28$ and $p_B = 28$, buyers hold beliefs $\mathit{q}(p', 28) = (q_A, 10)$. For prices $p_A = p_B = 28$, or for $p_A \neq 28$ and $p_B \neq 28$, buyers hold beliefs $\mathit{q}(p_A, p_B) = (q_A, 1)$.

The incentives to stick to this profile are very similar to those in cases (i) and (ii) in the Ability treatments without an entry barrier and are therefore omitted.

Case (iii) (no seller invested, both qualities are not observable for buyers): The possible equilibrium outcomes are the same as in the case in which both sellers invested in the Ability treatments.

Entry barrier

If the investment is a prerequisite to enter the market in the first place, the profits in the three scenarios are as follows:

Case (i) (both sellers invested, both qualities are observable for buyers): The possible equilibrium outcomes are the same as in the case with no barrier.

Case (ii) (only one seller invested, its quality is observable by buyers): The single seller that invested (seller A) becomes a monopolist. For a given quality, the monopolist can either serve all buyers in the market or cater to only a share of them. In the first case, the monopolist charges an optimal price of

$$20 + q - p = 0 \Leftrightarrow p = 20 + q.$$

The resulting profit is $(20 + q - q) \cdot 4 = 80$ (gross of investment costs), which is independent of the quality level chosen.

If the monopolist does not serve the whole market, the buyer who is indifferen between buying from the monopolist and not buying is given as

$$20 + \tilde{\theta} q - p = 0 \Leftrightarrow \tilde{\theta} = \frac{p - 20}{q}.$$

From the monopolist's maximization problem

$$\max_p \pi = (p - q)\frac{4}{3}\left(4 - \frac{p - 20}{q}\right),$$

the optimal price and the resulting profit can be derived as

$$p = \frac{5(4 + q)}{2} \qquad \text{and} \qquad \pi = \frac{(20 + 3q)^2}{3q}.$$

The indifferent buyer is then given by $\tilde{\theta} = 5(q-4)/2q$. Since we assumed $5(q-4)/2q \geq 1$, it must hold that $q \geq 7$. Also note that $\partial\pi/\partial q > 0$ for $q \geq 7$. Thus, the monopolist sets the maximum quality $q = 10$, which implies a price $p = 35$ and a profit $\pi = 83.\bar{3}$ (gross of investment costs). Thus, this is the more profitable option for the monopolist.

Case (iii) (no seller invested, both qualities are not observable for buyers): Both sellers are not active in the market and therefore make zero profit.

2.A.2.2 Investment decisions

Given the equilibrium in the subgames, we can determine the optimal (sequential) investment decisions of the sellers.

No entry barrier

Seller A has a first-mover advantage when it comes to investing in qualification. The investment allows the seller to inform buyers about its high quality, which allows the seller to earn higher profits than its competitor in the quality-then-price subgame. Considering the investment cost of 16, the higher profits in the quality-then-price subgame justify the initial investment. The predicted equilibrium strategies for sellers along the equilibrium path are as follows:

- $invest_A = 1$, $q_A = 10$, $p_A(p_B) = 28$ and $invest_j = 0$, $q_j = 1$, $p_j(p_i) = 10$ if buyer beliefs are as formulated in case (ii) above..

Entry barrier

Sellers never both invest, because the profit for the low-quality seller in the quality-then-price subgame of 12 is less than the initial investment. As a result, seller A becomes a monopolist (first-mover advantage) and serves part of the market. The predicted equilibrium strategies for sellers along the equilibrium path are as follows:

- $invest_A = 1$, $q_A = 10$, $p_A = 35$, and $invest_B = 0$ (monopoly).

Comparison

As discussed, maximum differentiation leads to an average quality that is lower than the monopolist's quality. However, prices are also lower in the case of differentiation, which leads to a higher consumer surplus in the case of no entry barrier.

In Table B.2, we summarize the predicted strategies. We include the demand for both sellers' products.

Table B.2: Equilibria TRANSPARENCY scenario.

	NOBARRIER		BARRIER	
Equilibrium	Differentiation		Monopoly	
Seller	A	B	A	B
Investment	1	0	1	0
Quality	10	1	10	-
Price	28	10	35	-
Demand	$\frac{8}{3}$	$\frac{4}{3}$	$\frac{10}{3}$	-
Profit	32	12	$83.\bar{3}$	-

2.B Robustness checks

2.B.1 Demographics

Table B.3 presents demographic statistics about our subjects in the experiment. We present the statistics for each scenario and separately by treatment. We find no significant differences between the treatments in both scenarios for any variable.

Given the fairly balanced demographics of our treatments, we should not see any changes in our results when we include them in the main regressions. To see if our results from Table 2.3 are robust when we control for different demographics, we run the same regression to compare treatments including the controls in Table B.4. For the Ability scenario, we see no significant changes in our estimated coefficients when we control for demographics. The same is true for the Transparency scenario. In Table B.5 we present the regression results including the controls. The estimated coefficients are also very similar to those in Table 2.4. Our main regression results are not driven by any demographic imbalances between our treatments.

Table B.3: Demographics.

	Ability			Transparency		
	both	A/NB	A/B	both	T/NB	T/B
female	0.62	0.65	0.58	0.54	0.53	0.56
	(0.157)	(0.178)	(0.139)	(0.179)	(0.146)	(0.222)
econ_stem_field	0.51	0.54	0.47	0.53	0.58	0.49
	(0.172)	(0.156)	(0.195)	(0.130)	(0.105)	(0.144)
experience	0.50	0.46	0.54	0.56	0.58	0.54
	(0.112)	(0.0697)	(0.137)	(0.151)	(0.190)	(0.115)
age	27.82	26.98	28.67	28.35	27.36	29.34
	(2.184)	(2.082)	(2.111)	(3.209)	(2.122)	(3.975)
risk	3.85	3.82	3.88	4.27	4.18	4.36
	(0.448)	(0.295)	(0.593)	(0.368)	(0.439)	(0.292)
trust	3.99	3.93	4.04	4.18	4.08	4.28
	(0.337)	(0.382)	(0.311)	(0.479)	(0.560)	(0.411)
trustme	5.72	5.78	5.67	5.85	5.76	5.93
	(0.217)	(0.209)	(0.230)	(0.369)	(0.359)	(0.392)
N	12	6	6	12	6	6

Notes: This table represents a summary of demographic statistics reported by the subjects in the experiment. *female* represents the share of subjects in the experiment that classified themselves as female when we asked them about their gender. *econ_stem_field* represents the proportion of subjects who reported to either study economics/business administration or a stem field. *experience* reports the proportion of subjects who have self-reported that they participated more than 10 times in experiments. *age* shows the average age of the subjects. *risk* asked whether subjects are a risk-taking person in general. *trust* asked: "How much do you trust other people?", and in *trustme* subjects report how much others can trust them. *risk*, *trust*, and *trustme* are all measured using a rating scale from 1 (not at all) to 7 (*risk*: fully risk-taking, *trust*: fully trust, *trustme* fully trust). Testing for equality of demographics between the treatments within a scenario, we find no significant difference in any variable using post-estimation Wald tests.

Table B.4: Regression analysis of the Ability scenario including controls.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	0.946***	0.499*	-0.281	-0.606	4.292	2.382	-22.10*	-20.17	-21.40***	-17.85***
	(0.118)	(0.242)	(0.349)	(0.549)	(2.234)	(2.763)	(9.471)	(11.93)	(3.191)	(4.678)
period		-0.0454**		-0.0691**		-0.527***		0.880**		0.309*
		(0.0149)		(0.0244)		(0.129)		(0.295)		(0.149)
barrier × period		0.0426**		0.0307		0.179		-0.184		-0.338
		(0.0151)		(0.0362)		(0.148)		(0.592)		(0.233)
Constant	0.931**	1.408***	2.665	3.367	11.08	16.40**	60.39***	51.15***	75.94***	72.69***
	(0.294)	(0.353)	(1.838)	(1.855)	(6.264)	(6.144)	(8.683)	(9.940)	(13.09)	(13.22)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.4636141	.5252105	.0628449	.0924032	.1758599	.2918489	.2170009	.2448104	.3620873	.3669662
Observations	940	940	913	913	913	913	1920	1920	940	940

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the Barrier treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. Included controls: *gender*, *field*, *experience*, *age*, *risk*, *trust*, and *trustme*. Details on these controls can be found in Table B.3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.5: Regression analysis of the Transparency scenario including controls.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	0.960***	0.938***	4.392***	3.405**	6.936**	6.670	16.50	10.76	4.063	0.530
	(0.152)	(0.179)	(0.602)	(1.297)	(2.568)	(3.686)	(12.25)	(17.36)	(4.638)	(9.342)
period		-0.0132		-0.0147		-0.322*		1.022		0.548
		(0.0121)		(0.0856)		(0.137)		(0.786)		(0.522)
barrier × period		0.00208		0.0956		0.0194		0.547		0.336
		(0.0132)		(0.0885)		(0.156)		(1.014)		(0.636)
Constant	0.458	0.596	2.393	2.570	2.554	5.846	126.7***	116.0***	106.6***	100.8***
	(0.386)	(0.451)	(2.454)	(2.849)	(8.366)	(8.466)	(20.13)	(21.91)	(15.83)	(15.99)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	.4133105	.4216918	.3779226	.3855416	.1898978	.2309278	.0968028	.1460074	.0318036	.058943
Observations	940	940	894	894	894	894	1900	1900	940	940

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the Barrier treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. Included controls: *gender*, *field*, *experience*, *age*, *risk*, *trust*, and *trustme*. Details on these controls can be found in Table B.3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.B.2 Subgame analysis

In this section, we complement the investment subgame analysis from section 2.7 by including also markets in which no differentiation occurred. Table B.6 shows the means and test results for the Ability scenario. For the Transparency scenario, the means and test results are presented in Table B.7. The results are in line with our findings in section 2.7.

Table B.6: Point predictions test Ability scenario (total).

NoBarrier

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	1	1	0	10	1	9	10	1	9
p-value	n/a	n/a	n/a	< 0.001	n/a	< 0.001	< 0.001	< 0.001	< 0.001
Mean	1	1	0	3.17	1	2.17	4.79	2.59	2.20
(Std. Dev.)	-	-	-	*(2.20)*	-	*(2.20)*	*(2.65)*	*(1.81)*	*(1.90)*
N (total)	75			114			51		

Barrier

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	-	-	-	1	-	-	10	1	9
p-value	n/a	n/a	n/a	$= 0.30$	n/a	n/a	< 0.001	$= 0.03$	< 0.001
Mean	-	-	-	1.63	-	-	2.26	1.28	0.98
(Std. Dev.)	-	-	-	*(1.28)*	-	-	*(1.96)*	*(0.85)*	*(1.72)*
N (total)	0			27			213		

Notes: We comprise the two investment combinations in which only one seller invests. In these subgames, we designate the investing seller as seller A in this table. In cases where both sellers have invested, we designate the one who offers the higher quality as seller A (if applicable). There are no cases when neither of the sellers invested in Barrier. Bootstrapped standard errors are clustered at the session level. We report p-values for post-estimation Wald tests. We test the Null Hypothesis (H_0): mean = prediction. For p-values below 0.05, we reject the null hypothesis, indicating that the observed outcome significantly deviates from the prediction. Pred notes the predicted value and Mean the experimental result. *N (total)* is the total number of markets in the respective investment subgame.

Table B.7: Point predictions test TRANSPARENCY scenario (total).

NOBARRIER

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	10	1	9	10	1	9	10	1	9
p-value	< 0.001	$= 0.18$	< 0.001	< 0.001	$= 0.015$	< 0.001	$= 0.02$	< 0.001	< 0.001
Mean	1.4	1.03	0.65	7.49	1.68	5.81	8.83	7.23	1.60
(Std. Dev.)	*(1.33)*	*(0.20)*	*(1.58)*	*(2.74)*	*(1.57)*	*(2.79)*	*(2.07)*	*(2.87)*	*(2.01)*
N (total)		83			116			41	

BARRIER

Investments	(0,0)			(1,0)/(0,1)			(1,1)		
	A	B	Δ	A	B	Δ	A	B	Δ
Pred	-	-	-	10	-	-	10	1	9
p-value	n/a	n/a	n/a	$= 0.23$	n/a	n/a	$= 0.02$	< 0.001	< 0.001
Mean	-	-	-	9.23	-	-	9.44	8.22	1.22
(Std. Dev.)	-	-	-	*(1.57)*	-	-	*(1.00)*	*(2.20)*	*(1.95)*
N (total)		6			35			199	

Notes: We comprise the two investment combinations in which only one seller invests. In these subgames, we designate the investing seller as seller A in this table. In cases where both sellers have invested, we designate the one who offers the higher quality as seller A (if applicable). Bootstrapped standard errors are clustered at the session level. We report p-values for post-estimation Wald tests. We test the Null Hypothesis (H_0): mean = prediction. For p-values below 0.05, we reject the null hypothesis, indicating that the observed outcome significantly deviates from the prediction. Pred notes the predicted value and Mean the experimental result. N *(total)* is the number of markets in the respective investment subgame.

2.B.3 Non-parametric tests

Table B.8 shows the results of non-parametric tests, comparing the two treatments within each scenario. The table reports medians and interquartile ranges (IQR) for the different variables. The ranksum tests are performed using the session means. The test results confirm our parametric test results, reported in section 2.6.

Table B.8: Comparison of treatments.

Scenario: Ability

Variable	NoBarrier	Barrier	Ranksum Test
	Median (IQR)	Median (IQR)	p-value
Investment	0.88 (0.35)	1.93 (0.1)	0.002
Quality	2.06 (1.44)	2.01 (0.79)	0.59
Price	10.48 (3.63)	14.83 (4.94)	0.13
CS	71.11 (20.97)	44.95 (29.10)	0.04
Welfare	79.26 (4.90)	59.25 (7.78)	0.002
N	6	6	

Scenario: Transparency

Variable	NoBarrier	Barrier	Ranksum Test
	Median (IQR)	Median (IQR)	p-value
Investment	0.83 (0.25)	1.93 (0.02)	0.002
Quality	3.71 (1.57)	8.14 (1.34)	0.002
Price	13.61 (10.10)	17.22 (4.67)	0.31
CS	88.85 (27.13)	96.13 (26.25)	0.39
Welfare	93.36 (15.00)	102.05 (4.42)	0.48
N	6	6	

Notes: We report the median as well as the Interquartile range (IQR $= Q3 - Q1$) in parentheses for the two treatments. We then report the p-value for the Wilcoxon ranksum test, also known as Mann-Whitney two-sample statistic (Wilcoxon, 1945; Mann and Whitney, 1947). This tests the hypothesis that two independent samples are from populations with the same distribution.

2.B.4 Excluding first rounds

Before the actual 20 periods of the experiment, we included two rounds of trials that are not included in any analyses in this paper. These rounds allowed subjects to become familiar with the post-instruction setting. To address the potential criticism that two rounds may have been too short a period, we now exclude the first two rounds of each session in our main regressions. This is a robustness check for our experimental results, to see if our findings in section 2.6 hold when we exclude rounds that may still have had a training character. As tables B.9 and B.10 show, our results are not substantially changed, as all hypothesis tests hold, even if we exclude these two rounds.

Table B.9: Regression analysis of the Ability scenario in rounds 3 – 20.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	1.032***	0.582*	-0.156	-0.478	4.559	3.802	-24.29*	-24.41	-22.21***	-20.99***
	(0.113)	(0.272)	(0.366)	(0.577)	(2.668)	(2.858)	(9.898)	(12.53)	(3.356)	(5.255)
period		-0.0416**		-0.0559*		-0.405***		0.799***		0.173
		(0.0161)		(0.0248)		(0.0956)		(0.227)		(0.110)
barrier × period		0.0391*		0.0279		0.0647		0.0110		-0.106
		(0.0163)		(0.0383)		(0.116)		(0.517)		(0.213)
Constant	0.852***	1.331***	2.137***	2.779***	11.65***	16.31***	63.36***	54.17***	79.70***	77.71***
	(0.0909)	(0.266)	(0.271)	(0.404)	(2.238)	(2.249)	(6.101)	(8.589)	(1.449)	(2.577)
Rounds	3–20	3–20	3–20	3–20	3–20	3–20	3–20	3–20		3–20
R^2	.4708517	.5122012	.00172	.0167512	.0896623	.1546461	.209159	.2338604	.3812551	.3826796
Observations	864	864	839	839	839	839	1728	1728	864	864

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the Barrier treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.10: Regression analysis of the TRANSPARENCY scenario in rounds 3 – 20.

	Investment		Quality		Price		CS		Welfare	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)
barrier	1.014***	1.132***	4.767***	4.327***	6.473*	7.268	12.81	7.231	4.843	-1.119
	(0.146)	(0.181)	(0.438)	(1.271)	(2.640)	(3.866)	(13.03)	(18.60)	(4.432)	(10.42)
period		0.00129		0.0293		-0.199		0.997		0.444
		(0.0112)		(0.0834)		(0.133)		(0.805)		(0.600)
barrier × period		-0.0102		0.0391		-0.0723		0.485		0.518
		(0.0121)		(0.0884)		(0.156)		(0.980)		(0.696)
Constant	0.773***	0.758***	3.917***	3.579**	13.36***	15.65***	77.86***	66.39***	95.88***	90.78***
	(0.0586)	(0.149)	(0.342)	(1.162)	(1.958)	(3.355)	(8.989)	(13.93)	(3.332)	(8.396)
Rounds	3–20	3–20	3–20	3–20	3–20	3–20	3–20	3–20	3–20	3–20
R^2	.4318049	.4336514	.3990691	.4041833	.1412532	.1617691	.0349245	.0714796	.0091301	.0326755
Observations	864	864	818	818	818	818	1728	1728	864	864

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *barrier* is the dummy-variable that takes the value of 1 for the BARRIER treatment. *period* reflects the time trend. *barrier* × *period* is the interaction of *barrier* and *period*. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.C Instructions

In Appendix 2.C, we provide the instructions for the online-laboratory experiment. Changes depending on the treatment are indicated as such using symbols. The legend for the symbols is as follows:

⋆ ***Ability/NoBarrier:***

⋄ ***Transparency/NoBarrier:***

∗ ***Ability/Barrier:***

⊚ ***Transparency/NoBarrier:***

Instructions for the experiment

Welcome to this experiment and thank you very much for your participation!

Some information about the general procedure of today's experiment:

- Below we will first show you the instructions for this experiment. Please read them carefully. All experiment participants will receive the same instructions. The written instructions will also always be available to you in the further course of the experiment via a button "Instructions". Please do not hesitate to contact the experimenters via chat if you have a question.

- After carefully reading the instructions, we will ask you comprehension questions. Once all participants have answered the questions correctly, the experiment will begin.

- First, a trial run of the experiment will start over two rounds to help you better understand how the decisions will be made. Winnings that you achieve in these rounds are not taken into account in the payout at the end of the experiment. Otherwise, however, there is no difference between the trial rounds and the later rounds.

- Then the actual experiment begins with a total of 15 rounds.

If you have any questions about the course of the experiment, please feel free to post them in the chat. Your questions will not be visible to the other participants.

Payments

In today's experiment, you will make repeated decisions in a market. You will play either the role of a seller or the role of a buyer in a market. More detailed information about these two roles will follow. Your payoffs at the end of the experiment depend on both your own decisions and those of the other participants.

During the experiment, all prices, costs and profits are quoted in ECU (experimental currency unit). Your payout will be anonymous and via PayPal at the end of the experiment. Your ECU score will be paid at a fixed rate of

$$50\ ECU = 1\ Euro.$$

At the beginning of the experiment, you receive a starting capital of 350 ECU (equivalent to 7 euros). This starting capital is available from the beginning of the experiment. Your account balance can therefore decrease or increase with your decisions during the experiment. After each round, we will show you your current account balance and your profit (or loss) from that round. You will also receive a fixed reward of 2.50 euros for your participation today.

Roles and market allocation

In this experiment, you will make decisions in a market as either a seller or a buyer. Which of these two roles you will take will be drawn at random at the beginning of the experiment. You will keep this role for the whole experiment. Your role will be told to you on the first screen when the experiment starts.

In a market, two sellers and four buyers meet in each round. The composition of a market is randomly determined each time.

The following diagram illustrates this division:

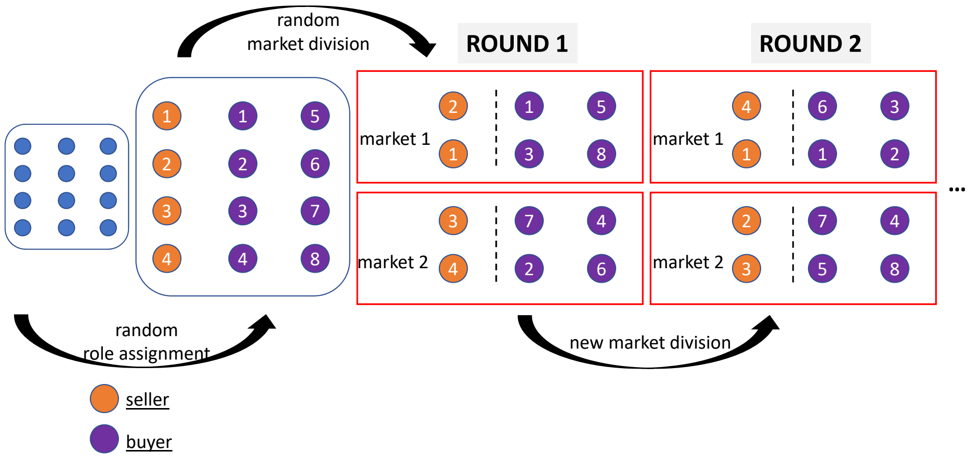
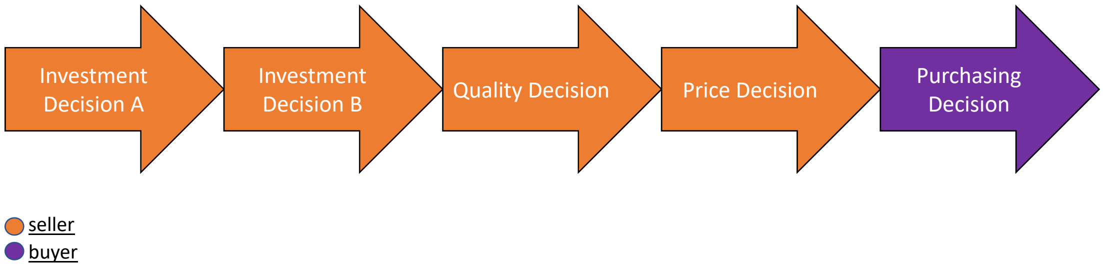


Figure B.1: Random role allocation before the start of the experiment into sellers (four in total, orange here) and buyers (eight in total, purple here). Then, in each round, two random sellers and four random buyers form a market.

Course of the rounds and decisions of the sellers and buyers



In each round, each of the two sellers can offer one product to each of the four buyers. The buyers can only buy a product from one of the sellers in each round or decide not to buy anything. However, before the buyers receive an offer, the sellers make decisions about their product.

At the beginning of each round, the sellers first decide whether they want to invest in their own qualification. This decision is made by the two sellers sequentially, with a draw in each round to determine who will make the investment decision first. The seller who makes the investment decision first is seller A, the other is seller B. Seller B thus knows the decision of seller A before he makes his own decision. The cost of an investment is 16 ECU.

The following part depends on the treatment:

⋆ ***Ability/NoBarrier:***

⋄ ***Transparency/NoBarrier:***

∗ ***Ability/Barrier:***

⊚ ***Transparency/NoBarrier:***

- ⋆ By investing in a qualification, a seller can select a quality for his product. The quality of the product can be chosen between values from 1 to 10. If the seller has decided not to invest in his qualification, his quality level is 1.

- ⋄ If a seller has decided to invest in a qualification, buyers can see the quality of the product offered by that seller before making their decision. If a seller has decided not to invest, the quality of the product offered is not apparent to buyers before they make their decision to buy. Sellers can choose between values from 1 to 10 for the quality of their product regardless of their investment decision.

- ∗ Only by investing in a qualification can a seller offer a product to buyers. If the seller has decided not to invest in his qualification, he cannot offer a product in this round. If a seller has invested in his qualification, he can choose a quality for his offered product. For the quality of the product, values from 1 to 10 can be chosen.

- ⊚ Only by investing in a qualification can a seller offer a product to buyers. If the seller has decided not to invest in his qualification, he cannot offer a product in this round. If a seller has invested in his qualification, he can choose a quality for his offered product. For the quality of the product, values from 1 to 10 can be chosen.

The following part depends <u>not</u> on the treatment:

Depending on the quality selected, additional costs are incurred by the seller, which are shown in the following table:

Quality	1	2	3	4	5	6	7	8	9	10
Production costs per-unit in ECU	1	2	3	4	5	6	7	8	9	10

In summary, if a seller decides to invest in a qualification, there is thus a cost of 16 ECU plus an additional cost of 1 to 10 ECU for production, depending on the quality chosen.

The following part depends on the treatment:

⋆ For this, the seller can choose a quality level of his product from 1 to 10.

If a seller decides not to invest in a qualification, he can only choose a quality level of 1. For this, there are then no investment costs and only the costs for production corresponding to the chosen quality 1, i.e. 1 ECU.

◇ In return, the quality level of his product, which he can choose from 1 to 10, is visible to buyers in advance.

If a seller decides not to invest in a qualification, the quality of his offered product, which he can choose from 1 to 10, is not visible to buyers. In return, there are no investment costs and only the costs of 1 to 10 ECU for the production, according to the chosen quality.

∗ For this, the seller can offer a product and choose a quality level of his product from 1 to 10.

If a seller decides against investing in a qualification, he cannot offer a product to the buyers in this round.

⊚ For this, the seller can offer a product and choose a quality level of his product from 1 to 10.

If a seller decides against investing in a qualification, he cannot offer a product to the buyers in this round.

The following part depends not on the treatment:

Then each seller chooses the price for which his product will be offered to each of the four buyers. Now it is the buyers' turn to make a decision. A buyer can choose between the two products offered at the price set by the respective seller and the alternative option of buying nothing. A buyer thus decides whether, and if so, which of the products offered should be bought or whether nothing should be bought.

The following part depends on the treatment:

- ⋆ If a seller has invested before, the buyer can only know the quality of the product after the purchase. If a seller has not invested before, the quality of the product is always 1. Thus, if the seller has not invested, the quality of the product is also known to the buyer before he buys it.

- ◇ If a seller has invested before, the buyer can see the quality of the product before buying. If a seller has not invested before, the quality of the product is not visible to the buyer.

- ∗ Every buyer can only recognise the quality of the product after purchase.

- ◎ Every buyer can see the quality of the product before making a purchase decision.

The following part depends not on the treatment:

A buyer can buy a maximum of one unit per round. A seller can sell one product to each of the four buyers. Before going into more detail about the influence of quality and the payoffs for buyers and sellers, here is a summary of the process and the decisions of the sellers and buyers in each round:

1. Two random sellers and four random buyers are brought together in a market for this round.

2. It is randomly determined which seller will make his investment decision first. This is seller A for this round.

3. Seller A makes his investment decision.

4. Seller B observes seller A's decision and then also makes his investment decision.

5. Seller A and B each know their own decision on the investment and also know what the other seller has decided.

The following part depends on the treatment:

⋆ 6. If a seller has decided to invest, he now chooses a quality of his product from 1 to 10.

7. Seller A and B each know their own quality and the quality of the other seller. The two sellers each choose a price for their product.

8. Each buyer can choose between seller A's product, seller B's product and the option to buy nothing. The quality of the products offered is only known to the buyer before making a decision if a seller has not invested and thus only offers quality 1.

9. Buyers and sellers learn their respective payouts as well as their account balances after this round. In addition to the quality, sellers also learn how many units the other seller has sold in this round and at what price.

⋄ 6. Each seller chooses a quality from 1 to 10 for his product. Provided a seller had invested, the quality of their product is apparent to buyers before they make a decision.

7. Seller A and B each know their own quality and the quality of the other seller. The two sellers each choose a price for their product.

8. Each buyer can choose between seller A's product, seller B's product and the option to buy nothing. The quality of the products offered is known to the buyer before the decision only if a seller had invested.

9. Buyers and sellers learn their respective payout as well as their account balance after this round. In addition to the quality, sellers also learn how many units the other seller has sold in this round and at what price.

∗ 6. If a seller has decided to invest, he will offer a product in this round and now chooses a quality of his product from 1 to 10.

7. Seller A and B know whether they themselves and also whether the other seller offer a product. They know the quality of their own product as well as the quality of the other seller if he offers a product. The sellers each choose a price for their product.

8. Each buyer can choose between seller A's product, seller B's product and the option to buy nothing. The quality of the products offered is not known to the buyer before making the decision.

9. Buyers and sellers learn their respective payout as well as their account balance after this round. In addition to the quality, sellers also learn how many units the other seller has sold in this round and at what price.

◎ 6. If a seller has decided to invest, he will offer a product in this round and now chooses a quality of his product from 1 to 10.

7. Seller A and B know whether they themselves and also whether the other seller offer a product. They know the quality of their own product as well as the quality of the other seller if he offers a product. The sellers each choose a price for their product.

8. Each buyer can choose between seller A's product, seller B's product and the option to buy nothing. The quality of the products offered is known to the buyer before the decision is made.

9. Buyers and sellers learn their respective payout as well as their account balance after this round. In addition to the quality, sellers also learn how many units the other seller has sold in this round and at what price.

The following part depends not on the treatment:

Calculation of the payout

The quality of a product affects both seller and buyer. As described earlier, the seller's costs increase with quality. Thus, the production of a higher quality causes higher costs for the seller than the production of a lower quality. The production costs for the different qualities are known to buyers and sellers and are shown again in this table:

Quality	1	2	3	4	5	6	7	8	9	10
Production costs per-unit in ECU	1	2	3	4	5	6	7	8	9	10

At the same time, a product with higher quality is worth more to buyers than a product with lower quality. The quality valuation of a buyer is randomly assigned to each buyer in each round. The individual valuation can take on any value between 1.00 and 4.00 with equal probability.

The amount of one's own valuation for quality (X) is only known to the respective buyer. Neither the other buyers nor the sellers know this.

For a buyer, the value of a product is given by

$$Value = 20 + (Quality \times X)$$

Accordingly, a product brings each buyer at least a fixed value of 20 plus the variable value resulting from the quality of the product and the buyer's individual valuation of quality.

The payoff for each buyer is then the difference between the value the product has for the buyer and the price the buyer paid for the product to the seller:

$$Payment = Value - Price$$

If a buyer decides not to buy anything, his payout in this round is 0 ECU.

For the seller, the quality of the product also has an influence on the payout, as does the investment decision made beforehand. In addition, the number of units sold plays a role. A seller could sell one product to each of the four buyers in the market.

If a seller decided to invest, the cost of doing so of 16 ECU is incurred regardless of whether and how many units were sold. The production costs for one unit are again only incurred for the units actually sold, but they vary according to quality. If a seller has invested, his payout is given by:

$$Payment = (Number\ of\ units\ sold) \times (Price - Production\ costs) - 16.$$

The following part depends on the treatment:

⋆ If a seller has not invested, his quality is 1 and his payoff is given by:

$$Payment = (Number\ of\ units\ sold) \times (Price - 1).$$

⋄ If a seller has not invested, his payout is given by:

$$Payment = (Number\ of\ units\ sold) \times (Price - Production\ costs).$$

∗ If a seller has not invested, his payout in this round is 0 ECU.

⊚ If a seller has not invested, his payout in this round is 0 ECU.

Chapter 3

Redistribution, Moral Hazard, and Voting by Feet: An Experiment

3.1 Introduction

Income inequality and how to reduce it with redistribution policies are high on the agenda of public debate. The World Inequality Report (Chancel et al., 2022) argues that there has been an increase in inequality in the 20th century, especially since the 1970s, and emphasizes the need for tax policies and wealth redistribution to address this problem. Stiglitz (2012) argues that inequality is not only a moral issue, but also a major economic problem that is harmful to economic growth, stability, and democracy. He provides a series of policy recommendations for addressing inequality, including progressive taxation. The OECD has reported (Sarfati, 2015) that income inequality has increased in many of its member countries and has emphasized the need for policies that promote inclusive growth.

Parallel to the public debate is a longstanding academic debate over the nature of redistributive preferences. To ensure unbiased statements about preferences and to avoid self-serving biases, researchers often refer to Rawls' veil of ignorance. Rawls (1971) proposed a thought experiment that asks us to imagine what kind of society we would choose if we knew nothing about our own identities, abilities, preferences, or circumstances. This idea has been taken up in recent laboratory experiments to investigate the redistributive preferences of participants. Related studies, reviewed below, include Frohlich et al. (1987), Frohlich and Oppenheimer (1990), Konow (2000), Andreoni and Miller (2002), Sutter and Weck-Hannemann (2003), Engelmann and Strobel (2004) Cappelen et al. (2007), Schildberg-Hörisch (2010), Durante et al. (2014), Deffains et al. (2016), and Gerber et al. (2019). The studies show (see section 3.2) that a substantial fraction of participants exhibit non-selfish preferences for redistribution. Certainly, there are subjects who choose opportunistically, and such behavior may be reinforced by self-serving biases. But preferences for redistribution, especially behind a veil of ignorance, appear to be substantial.

Given the experimental evidence, the question arises of whether and how these preferences for redistribution materialize in actual tax policies. One channel through which citizens can express their preferences and affect taxes is their vote in elections. While the impact of an individual vote in a democratic election is limited, as decision-makers follow the median voter (Meltzer and Richard, 1981), with increasing inequality, support for redistribution may increase and find more voters. If so, the demand for redistributive policies should increase, and redistribution would rise.[1]

However, even assuming that redistribution policies find significant support and are implemented with progressive income taxes, there are two challenges: First, once the principles of a fair tax system was established, such a system would be subject to competition from other systems in other countries or jurisdictions ("voting by feet"). Given the heterogeneity of systems and the fact that some agents act opportunistically, voting by feet would put such a redistributive system under pressure. Tiebout (1956) argues that individuals can express their preferences for public goods and services (and hence demands on the tax system) by moving from one jurisdiction to another, ideally to the one that offers the combination of public goods, taxes, and services that best match their preferences. Next to the question of what constitutes a fair income distribution, a policy-relevant question is how sustainable is such a system in national and international tax competition?[2] This is also related to the theoretical predictions of unraveling theory (Nagel, 1995), which suggests that people update beliefs and adjust behavior based on individual experience. Following the logic of unraveling, people choose the jurisdiction that best serves their preferences after information is available that allows them to update their beliefs.

A second challenge are the inefficiencies that arise due to redistribution. Economists have long emphasized that there are moral-hazard effects when imposing (linear)

[1] It seems that, at least in the US, many people vote against redistributive policies such as more progressive income taxes and more transfers, even though they would benefit from them (see Bartels, 2008; Stantcheva, 2021).

[2] There is also a growing empirical literature on migration responses to income taxation, which finds that high-income and high-skilled taxpayers in particular are highly mobile. Kleven et al. (2020) provide a good overview of studies on both international and within-country mobility. Despite this, they also call for caution in terms of external validity and generalization, as the available data is limited and most studies rely on data for specific groups, such as the top 5 percent of earners, celebrities or entrepreneurs. Also Engelmann et al. (2023) provide some evidence from a survey experiment on the migration of high-income individuals and how political attitudes affect the decision to migrate.

income taxes. Okun (1975) coined the "leaky bucket" and attributed losses from redistribution not least to incentive effects: The less affluent have less incentive to work because their transfer payments are reduced if they make more money whereas the more affluent have less incentive to work because high marginal tax rates take a large fraction of their additional income.

The experimental literature cited above (and discussed in detail below) largely analyzes participants' fairness preferences for how to divide a given pie. Oftentimes, one subject must produce the pie, or the pie is exogenously given. A second subject decides how to divide the pie, as in a dictator game. In these settings, there is no competition from other tax systems, and since these are typically one-shot experiments there is presumably little moral hazard: fatigue may not have set in, and decisions may not yet have converged.

My paper adds to the literature by emphasizing that, first, tax systems are due to competition via voting by feet, and second, the moral-hazard problem arises because citizens work less hard when income is subject to a tax. I let subjects choose between two tax rates: The tax revenues are used for redistribution only (no inefficiencies or other tax-related costs, no government expenditures). The low rate implies very little redistribution, and the high rate leads to substantially more equal incomes. Subjects' decisions are choices (not votes) and thus match the decision as to whether to move to a jurisdiction with lower income taxation. After choosing the tax, subjects produce income from conducting a three minute real-effort task. The tax choice and the three-minute work phase are repeated six times, such that fatigue may reinforce moral hazard and tax choices may fully unravel.

Using this design, I address the following research questions.[3] First, what tax rate will subjects choose, and how do tax choices change over time? I distinguish between the initial choice under a Rawlsian veil of ignorance (subjects do not know the nature of the task or how well they will do in it) and later choices, which may

[3]The method of a laboratory experiment is well suited to exploring this topic and to addressing the research questions I want to answer in an appropriate way. The controlled environment allows both moral hazard and selection to be observed in isolation from other factors that might intervene in real life.

be driven by opportunistic rather than insurance motives. Later tax choices are likely to be driven by ability (income). My second question is: what are the effort (income) differences between the two tax regimes (over time) and do they depend on self-selection? To quantify the moral hazard effect and control for selection, I use a treatment with exogenous (random) tax rates. There is moral hazard due to the tax in both treatments, but with exogenous taxes, agents with different incomes are equally likely to end up with the high/low tax. A third treatment is suitable for eliciting subjects' attitudes toward income uncertainty. This treatment helps to identify whether unraveling affects tax choices. In this treatment, subjects still choose the level of the redistributive tax rate in the first stage, but their income is random in the second stage. Thus, there is uncertainty about income and the choice of tax rate reflects this income uncertainty.

The paper is organized as follows. Section 3.2 summarizes important literature on redistributional preferences and voting, which I relate to. In section 3.3, I present the theoretical background. The experimental design, theory-based hypotheses which have been pre-registered,[4] and procedure follow in sections 3.4, 3.5, and 3.6. Section 3.7 reports the results, before I conclude in section 3.8.

3.2 Literature

The experimental literature on redistributive preferences, inequality, and voting is rich. Many point to Rawls' veil of ignorance as a key element in eliciting true redistributive preferences.[5] Frohlich et al. (1987) brought the veil of ignorance into the laboratory.

[4]See https://doi.org/10.17605/OSF.IO/2CJKX.

[5]In "A Theory of Justice," Rawls (1971) uses this thought experiment to argue for a conception of justice based on the principles of fairness and equality. He advocates a redistributive system in which the distribution of resources and opportunities is arranged to benefit the least advantaged members of society (maximin rule). In contrast to Rawls, Nozick (1974), emphasizes individual liberty and the protection of property rights over redistribution. Harsanyi (1975) represents the utilitarian counterpoint to Rawls: a rational decision rule must be based on the expected utility of any institutional arrangement, and (behind a veil of ignorance) should use the average utility of individuals instead of the maximin rule, under the assumption of the equal likelihood of all possible outcomes. Last but not least, Buchanan (1976) proposes a theory of constitutional economics, arguing that a limited government that protects individual liberty and property rights is a more just system than one that prioritizes redistribution.

They test whether the Rawlsian maximin principle is the preferred distributive scheme when subjects choose among a set of different principles before being randomly assigned their income. Although the subjects preferred average maximization to the Rawlsian principle, they chose a floor to protect the lower income subjects. Frohlich and Oppenheimer (1990) build on these results and compare provided effort in a real-effort task between treatments in which distributive principles were either chosen (voting) or imposed. They find that when the distributive principle is exogenously imposed, subjects who receive transfers produce less effort than those who actively participate in the decision process.

In addition to the role played by active participation in the choice of a distributive scheme, risk attitudes and other-regarding preferences may also influence redistributive preferences. Konow (2000) studies distributional preferences in a dictator game with production to elicit the role of fairness considerations. In a two-player experiment, subjects produce income using a real-effort task. In contrast to Frohlich and Oppenheimer (1990), the distribution between pairs' joint income is decided by either an in-group dictator or a third-party dictator. Konow (2000) finds that third-party dictators with no financial incentive in the distribution between the two subjects distribute payoffs more equally when differences in contributions depend on luck rather than on subjects' effort. Also, Cappelen et al. (2007) add to this literature on the perceptions of fairness in distributive justice by testing for different fairness ideals. They test for distributive principles such as egalitarian fairness (equal share of total output for each person), libertarian fairness (each person gets exactly what she contributed), and liberal egalitarianism (each person gets the same share of total output as she invested). They find support for all three principles, with a majority favoring strict egalitarianism and equality of opportunity (liberal egalitarianism).

Engelmann and Strobel (2004) find that in simple distribution experiments not all of the subjects' decisions can be related to inequality aversion, but find that selfishness, concern for efficiency, and maximin preferences can explain most of it.[6] This result

[6]Engelmann and Strobel (2004) analyze different inequality aversion theories (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002) in different one-shot distribution

is related to Schildberg-Hörisch (2010) who also finds that efficiency matters for impartial decision-makers. Further, she finds that other-regarding preferences and risk aversion play a role in behind-the-veil decisions, such that many subjects prefer more equal allocations behind the veil of ignorance, while only a small fraction follow the Rawlsian maximin principle.

Deffains et al. (2016) study the demand for redistribution in combination with a real-effort task to assess the self-serving bias of subjects' redistribution decisions. By exogenously manipulating subjects' perceptions of success or failure in the task, Deffains et al. (2016) find that successful subjects link their success more to their effort rather than luck, while unsuccessful subjects blame external factors for their failure. In support of the self-serving bias hypothesis, they find that subjects favor more/less redistribution accordingly.[7]

Sutter and Weck-Hannemann (2003) directly link effort provision and taxation. In a two-person bargaining game, they study the tax choices made by one subject relative to the effort choices made by the other subject, with and without a veil of ignorance. They find that increasing taxes discourage effort.[8] Sutter and Weck-Hannemann (2003) find increasing taxes to be especially discouraging when they exceed 50 percent. They further show that this effort is less pronounced behind the veil of ignorance when subjects care more about the whole group.[9] Sutter and

experiments. For example, the authors study a taxation game in which subjects choose between different degrees of redistribution, like different tax systems. They then test which theories fit best and which additional effects rationalize their data.

[7]In a follow-up project, Espinosa et al. (2020) test whether information and the timing of this information provision changes the results for self-serving bias. While ex-post information does not change the self-serving bias, they find that information provision ex-ante prevents the formation of the self-serving bias.

[8]In a very recent working paper, also Mill and Schneider (2023) show the highly discouraging effect of a higher tax rate on labor supply. For theoretical discussions and supporting empirical evidence from natural experiments, see also Dalamagas and Kotsios (2008) and the references therein.

[9]Similar to Sutter and Weck-Hannemann (2003), Durante et al. (2014) also combine tax choices and income to study redistributive preferences. They also adopt various fairness ideals as in aforementioned studies. Durante et al. (2014) assign subjects income proportionally to the pre-tax distribution in the U.S. based on either random selection, socioeconomic background, or two skill-based tasks. By having subjects decide on taxes for each of the four assignment methods with and without a veil of ignorance, the authors explore redistributive preferences as well as social fairness concerns. Taking into account the efficiency loss of redistribution, they find that the demand for redistribution depends not only on inequality and efficiency concerns, but also on self-interest and insurance motives. This is also found in other studies, such as Andreoni and

Weck-Hannemann (2003) also include a voting decision before the second round. While this is mainly introduced to create uncertainty, other experiments focus more on actual voting decisions regarding redistribution and the size of the tax rate.[10]

My focus is on individual choices between two exogenous tax regimes, that is, on voting by feet. Redistribution and voting by feet in a laboratory experiment is also studied by Gerber et al. (2019). They combine the choice for a distributive principle with a voting by feet decision.[11] Unlike my experiment, they do not use a real-effort task to define the individual's contribution to the pie. Like mine, their experiment also involves repeated voting by feet decisions.[12] In contrast to my experiment, individual predispositions (the productivity factor) change in every round.[13] Gerber et al. (2019) find some support for all distributive principles in all treatments. Also, many subjects change their choice of distributive rule over the course of the 32 rounds, as few choose the same rule in more than 75 percent of the periods. They also find that the amount of redistribution decreases as subjects' knowledge of their productivity increase.

All of the above papers study redistributive preferences in different ways. My paper differs in the following key aspects: First, while some papers focus on the choice between different distributive principles, my paper implements the choice between a high or a low level of redistribution. Second, the pie in my experiment is generated by a real-effort task, while many others use, for example, windfalls and investments, or rank-based income distribution. Third, I use a voting by feet mechanism to let

Miller (2002), which find heterogeneity in fairness and inequality aversion to serve the full range of preferences from utilitarian to pure selfishness.

[10]Examples of redistribution experiments that include a (democratic) voting stage that is resolved by majority rule are Tyran and Sausgruber (2006), Höchtl et al. (2012), Cabrales et al. (2012), Jiménez-Jiménez et al. (2020).

[11]Voting by feet has also been studied in laboratory experiments related to institutional choices concerning e.g., social dilemmas (Gürerk et al., 2006; Gürerk et al., 2014), and in a broader sense as self-selection into different groups (see, e.g., Jaworski and Wilson, 2013).

[12]Gerber et al. (2019) compare the probabilities of choosing different redistribution principles (libertarian, proportional, and egalitarian) under different information treatments that are either no, full, or partial information.

[13]In Gerber et al. (2019), the subjects' contribution to the pie is the return on an investment, which they decide on after learning about an individual productivity factor. These factors, which are either 1.2, 2, or 4, are renewed each round. The distributional decisions are made with either full, partial, or no information about the individual factor.

subjects individually choose the amount of redistribution rather than, for example, dictatorial choices, median voting, or randomness. Fourth, I study repeated choices, which, in contrast to one-shot experiments, allow for the unravelling of tax choices and fatigue in the real-effort task.

3.3 Model

3.3.1 Basic setup

Agent i invests effort e_i and is paid a piece rate, r, such that she produces a taxable gross income $\pi^g = e_i r$. Effort is subject to a convex cost function, $C(e_i)/\theta_i$, with $C(e_i) \geq 0, C'(e_i) \geq 0, c''(e_i) > 0$ where θ_i is the player type. Her gross utility before taxes is:

$$u_i^g = \pi_i^g - C(e_i)/\theta_i = e_i r - C(e_i)/\theta_i.$$

Now consider a linear income tax which is purely redistributive: All tax payments are returned in equal shares to the agents. Let $t \in [0, 1]$ be the tax rate, $T_i = tre_i$ is i's tax payment (effort does not reduce the taxable income), and $\sum_{j=1}^{n} T_j/n$ is the transfer i receives. Then, i's net utility after tax payment and transfer is:

$$\begin{aligned} u_i^n &= u_i^g - T_i + \frac{\sum_{j=1}^{n} T_j}{n} \\ &= \pi_i^g \ (1-t) - \frac{C(e_i)}{\theta_i} + \frac{t}{n} \sum_{j=1}^{n} \pi_j^g \\ &= e_i r \ (1 - t + \frac{t}{n}) - \frac{C(e_i)}{\theta_i} + \frac{t}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} e_j r \end{aligned} \tag{3.1}$$

The effective tax, $1-t+\frac{t}{n}$, depends on the group size, although that effect is negligible in large groups. From (3.1), it follows that any agent with an above-average effort, $e_i > \sum_{j=1}^{n} e_j/n$, is a *net payer* (loses from redistribution) whereas any agent with a below-average effort $e_i < \sum_{j=1}^{n} e_j/n$ is a *net recipient*.

3.3.2 The moral-hazard effect of redistribution

The first-order condition of agent i reads

$$\frac{\partial u_i^n}{\partial e_i} = r(1 - t + \frac{t}{n}) - C'(e_i)/\theta_i = 0.$$

That is, the higher the reward r and the higher θ_i, the more effort results. Due to $0 \leq 1 - t + t/n \leq 1$, the optimal effort level is positive for any $t > 0$ and $n > 1$, but lower than without the tax ($t = 0$). This is the moral hazard effect of the tax. In a large group and with a tax rate of $t = 100\%$, all efforts would be zero.

3.3.3 Tax decisions behind a veil of ignorance

When deciding initially without knowledge of the task and their ability, subjects decide behind a veil of ignorance. This is decision-making under uncertainty where redistribution has an insurance effect, resembling the setup of Harsanyi (1975).[14]

Assumption 1 *There are two types of agents, $\bar{\theta}$ and $\underline{\theta}$, where $\bar{\theta} \geq \underline{\theta}$.*

Imposing Assumption 1 for simplicity, it turns out that $\bar{\theta}$ is a net payer and $\underline{\theta}$ is a net recipient. Assume that m of the n agents are high types ($\bar{\theta}$) and $n - m$ are low types ($\underline{\theta}$). In a large population, let $\rho = m/n$ be the prior for being the high type and $1 - \rho = (n - m)/n$ is the probability for the low type.

In order to isolate the redistribution aspect of the tax, I first leave aside its effort-reducing effect. Instead, I assume that agents have available some fixed gross income, regardless of t. I denote the two levels of gross income by $\bar{\pi}^g$ and $\underline{\pi}^g$. The net income, denoted by π^n, after taxes for the high type reads

$$\bar{\pi}_i^n = \bar{\pi}_i^g(1 - t) + t(\rho\bar{\pi}_i^g + (1 - \rho)\underline{\pi}_i^g) = \bar{\pi}_i^g - t(1 - \rho)(\bar{\pi}_i^g - \underline{\pi}_i^g)$$

[14] I leave social preferences aside here as the purpose of this theory section is that even rational (selfish) agents may or may not choose the high tax. An inequality-averse agent would be more inclined toward the high tax, but this would not change the conclusion that both tax rates may be chosen (see Hypothesis 1).

and, for the low type, it is

$$\underline{\pi}^n = \underline{\pi}_i^g(1-t) + t(\rho\bar{\pi}_i^g + (1-\rho)\underline{\pi}_i^g) = \underline{\pi}_i^g + t\rho(\bar{\pi}_i^g - \underline{\pi}_i^g).$$

Clearly, $t = 0$ implies that agents receive $\pi_i^n = \pi_i^g$ whereas $t = 1$ implies that both agents obtain the same net income, namely $\bar{\pi}_i^n = \underline{\pi}_i^n = \underline{\pi}_i^g + \rho(\bar{\pi}_i^g - \underline{\pi}_i^g)|_{t=1}$. Solving $\bar{\pi}^n = \bar{\pi}_i^g - t(1-\rho)(\bar{\pi}_i^g - \underline{\pi}_i^g)$ for t and substituting, I obtain the budget line

$$\underline{\pi}^n = \underline{\theta} + \frac{\rho}{(1-\rho)}\left(\bar{\theta} - \bar{\pi}^n\right),$$

indicating that the budget line has a slope of $-\rho/(1-\rho)$, a ratio corresponding to the relative frequency of types.

I now add the moral hazard aspect, for the sake of concreteness, with a parametrized cost function. Assume $C(e_i) = e_i^2/2\theta_i$ and simplify $r = 1$. For a large n, I obtain

$$u_i^n = e_i(1-t) - \frac{e_i^2}{2\theta_i}$$

and the optimal effort level reads $e_i = \theta_i(1-t)$. Due to $r = 1$, these efforts imply gross incomes of $\bar{\pi}_i^g = (1-t)\bar{\theta}$ and $\underline{\pi}_i^g = (1-t)\underline{\theta}$. Substituting yields

$$\bar{\pi}^n = (1-t)[\bar{\theta} - t(1-\rho)(\bar{\theta} - \underline{\theta})]$$

and

$$\underline{\pi}^n = (1-t)[\underline{\theta} + t\rho(\bar{\theta} - \underline{\theta})]$$

Unsurprisingly, the high type is always worse off with taxation. But even the low type is not always better off: This will be the case only if the income gap, $\bar{\theta} - \underline{\theta}$, is sufficiently large.

Figure 3.1 shows the effects of the redistribution tax for a risk-averse decision-maker: The point $(\bar{\pi}^g, \underline{\pi}^g)$ shows the gross income of the two types when $t = 0$. The downward-sloping dotted line illustrates the pure insurance effect if there is no moral hazard and where $(\bar{\pi}^e, \underline{\pi}^e)$ occurs when $t = 1$. The upward-sloping dashed line highlights the moral-hazard effect when there is no redistribution and where $(0, 0)$ results when $t = 1$. The concave solid, likewise ending in $(0, 0)$ when $t = 1$, takes both effects into account.

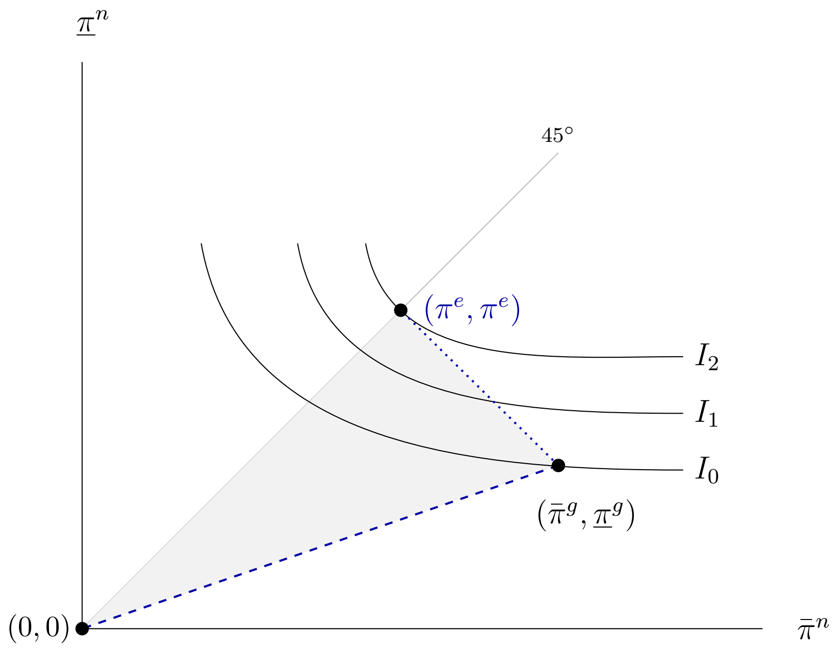


Figure 3.1: Decision-making under uncertainty: (i) dotted line reflects redistribution only without moral hazard, (ii) the dashed line shows moral hazard without redistribution, (iii) with both moral hazard and redistribution any point in the gray area is possible. Which point is realized, depends on the agent's risk attitudes and the effort cost.

There are several conclusions that can be drawn from Figure 3.1. First, a risk-averse agent would prefer $(\bar{\pi}^e, \underline{\pi}^e)$ to $(\bar{\pi}^g, \underline{\pi}^g)$. Thus, if there was no moral hazard, risk-averse agents would prefer $t = 1$. This is the insurance effect of redistribution. Second, assume hypothetically there is no redistribution. Then agents would prefer $t = 0$ as any point on the dashed line makes both types worse off. Third, taking both insurance and moral-hazard effect into account, risk-averse agents may or may not prefer a positive tax rate. This depends on their risk attitudes and the effort cost. In Figure 3.1, an improvement due to $t > 0$ is possible, but this is not a general conclusion. In other words, for some tax rate $\in (0, 1)$, risk-averse agents would prefer taxation to no taxation. But there is no unambiguous conclusion here as the reference for taxation depends on the risk preferences. Furthermore (and intuitively), agents prefer zero taxation the smaller the gap between $\bar{\pi}^g$ and $\underline{\pi}^g$: The gain from redistribution is small, but the moral-hazard effects still has full force.

3.3.4 Tax decisions knowing your type

After the first round of play, subjects know their type. This means that they know whether they are a net payer or net receiver, and the insurance effect no longer plays any role. Net payers ($\bar{\theta}$) are better off with the low tax rate and will as thus choose.If they choose to not respond, all net receivers ($\underline{\theta}$) would be left alone in the high tax regime and would thus not gain from redistribution. Anticipating that high types choose $\underline{t}$, all low types also choose the low tax regime. In theory, tax choices *unravel*: all agents choose the low taxes from round two on, assuming common knowledge of rationality.

The logic of immediate unraveling is based on a rational agent who perfectly anticipates the behavior of others. Limited depth of reasoning or level-k thinking (Nagel, 1995) instead suggests that boundedly rational agents might not perfectly anticipate the behavior of others. Subjects may choose the tax rate myopically: While net payers opportunistically choose the low rate (without the need to consider others), the net receivers may naively choose the high tax rate. Net receivers expect to benefit from the high tax without realizing that high types choose the low tax. After the second round of play, they realize their mistaken belief and may switch to the low regime themselves.[15] With more than two heterogeneous types, this implies that fewer and fewer agents choose the high rate. Over the course of the six repetitions, fewer and fewer subjects will choose the low rate.

[15]This process is analogous to the unraveling in the repeated beauty contest game described in Nagel (1995), where subjects learn to play the solution over four repetitions.

3.4 Experimental design

In my Baseline treatment, participants play six rounds. In each round, they first choose between a high and a low tax rate (voting by the feet). Second, they have to conduct a real-effort task. The income generated is then redistributed among those who chose this tax rate. I now describe these stages in detail.

Subjects choose between a high and a low tax rate before doing the real-effort task. The high rate is $\bar{t} = 85\%$ and the low rate is $\underline{t} = 15\%$. I emphasize in the instructions (and check the understanding with control questions) that the tax is entirely for redistribution: All tax revenues are returned to the agents in equal shares. There are no immediate inefficiencies, no direct costs of the tax system, and no other government expenditures. The only cost of the redistribution would be due to moral hazard.

Crucially, the first choice of the tax rate is behind a *veil of ignorance*. Subjects do not know the nature of the real-effort task and thus they are unaware of how good they are at the task. This distinguishes the choice in the first round from later rounds when subjects are familiar with the task and when they know their relative performance from the redistribution stage.

After subjects have made their tax rate choices, they are told how many other agents are in the high/low-tax regime. That is, they knew about the group size. Group sizes with just one or even no agent are possible.

Subjects then have to earn their income with a real-effort task. Specifically, subjects had to do a three minute decoding task.[16] They were paid 10 ECUs for each puzzle they solved. They knew immediately from the screen whether they had solved a puzzle correctly.

After the three minutes were over, participants received feedback. I told them how many puzzles they had solved correctly, how much gross income they had earned, their tax payments, and the transfer they received as part of the redistribution. The

[16]The real-effort task used in this experiment is similar to Benndorf et al. (2019). There were five digits that had to be translated into five letters using a decoding sheet that changed with each puzzle, as did the digits and letters. Like the whole experiment, this task was computerized.

Table 3.1: Treatment design.

		Tax choice	
		endogenous	exogenous
Income	endogenous	Baseline	Tax_Random
	exogenous	Income_Random	✗

feedback also included the average of each group's solved puzzles, transfers, and the payoff for both tax groups. Then the next round started.

As for experimental treatments, I conducted the following variants, see Table 3.1. The above-described Baseline treatment involves both the choice of tax rate and the real-effort (or income) choice. Since both decisions are mutually dependent (or endogenous), I compare Baseline to treatments where either the tax rate or the income is randomly varied and hence exogenous: In each round of Tax_Random, m participants are randomly assigned the high tax regime and the remaining $n - m$ agents have the low tax rate. Subjects learn about their assigned tax rate prior to their income-generating task. The proportion of subjects assigned to each tax group was taken from Baseline to ensure the comparability of treatments. Treatment Tax_Random thus controls for the selection effects of tax choice (high types are more likely to choose the low tax). In treatment Income_Random, income is randomly drawn from the actual distribution of gross incomes earned in Baseline. Since the tax is chosen before subjects learn their income, this treatment allows for conclusions about subjects' insurance motives. In each round, they are uncertain about their income position prior to the redistribution.

3.5 Hypotheses

In this section, I summarize the above theory approaches with several testable and pre-registered (https://doi.org/10.17605/OSF.IO/2CJKX) hypotheses. As is the case in the experiment, I now assume that subjects choose between two tax rates, a high tax rate and a low tax rate:

Assumption 2 *Agents can choose between two levels of tax* $\bar{t}$ *and* $\underline{t}$*, where* $\bar{t} \geq \underline{t}$.

I start with a hypothesis about the tax choice in the first round. This follows from subsection 3.3.3 and Figure 3.1.

Hypothesis 1 (H1) *(Null) When deciding behind a veil of ignorance, subjects may choose* $\bar{t}$ *or* $\underline{t}$.

The following set of hypotheses is about tax choices after the first round of the game. The first hypothesis is very rigorous and builds on the rational agent premise, following the unraveling logic in subsection 3.3.4. The second hypothesis tests the voting by feet mechanism over the course of the experiment, which is consistent with the boundedly rational agent framework. The third hypothesis also follows from the boundedly rational agent assumption.

Hypothesis 2a (H2a) *(Immediate unraveling) After the first round of play, all subjects choose* $\underline{t}$.

Hypothesis 2b (H2b) *(Gradual unraveling) Over the course of the experiment, the share of subjects choosing* $\underline{t}$ *increases.*

Hypothesis 2c (H2c) *After the first round of play, net payers choose* $\underline{t}$*, net receivers choose* $\bar{t}$.

Further, I test for a diminishing insurance effect and performance belief updating in Baseline compared to Income_Random. My directed hypothesis is

Hypothesis 3 (H3) *The time trend of the share of subjects choosing* $\underline{t}$ *is less pronounced in Income_Random compared to Baseline.*

Three more hypotheses are about effort provision. From the first-order condition of equation 3.1, I conclude the following hypothesis for the experiment concerning effort provision and tax choice.

Hypothesis 4 (H4) *The higher the tax rate, t, the lower the effort levels.*

The next hypothesis is a treatment comparison where I test for a selection effect in BASELINE compared to TAX_RANDOM. (The treatment description of TAX_RANDOM can be found in section 3.4.)

Hypothesis 5 (H5) *The effort difference between the subgroups high/low is more pronounced in BASELINE compared to TAX_RANDOM.*

The next hypothesis also follows from the first-order condition of equation 3.1. This hypothesis takes into account the finite group size in a laboratory experiment and relates effort provision to group size. Since population size, n, has an effect in theory, I hypothesize that larger groups exacerbate the moral hazard problem.

Hypothesis 6 (H6) *The larger the group size, n, the lower the effort levels.*

The last hypothesis considers behavioral factors for the tax decision. From the measure of confidence after the first round and the measure of risk I elicit, and based on Figure 3.1, I hypothesize:

Hypothesis 7 (H7) *The more confident and the more risk-loving subjects are, the more frequently they choose $\underline{t}$.*

3.6 Procedures

The experiments were conducted as laboratory sessions at the DICELab of Duesseldorf University and Cologne Laboratory for Economic Research (CLER) in June and July 2023. Subjects were recruited from existing pools of volunteers. At DICELab the database is based on and the recruiting was done in hroot (Bock et al., 2014). At CLER the recruiting was based on ORSEE (Greiner, 2015). The experiments are programmed in oTree (Chen et al., 2016).

The protocol was the same for all sessions. At the beginning of a session, participants read the instructions, which tell them about the experiment as a whole but without details on the income-generating process. Experimental instructions were shown on screen, but paper copies were also available at all times. The instructions were accessible to the participants during the whole course of the experiment. After reading these instructions, the participants had to answer several control questions before the first tax decision was made.[17] After the first round tax decision, subjects received detailed instructions for the income-generating process. From this point on, these were also always accessible for the participants. After the main experiment, I elicited measures for risk preferences, altruism, confidence, and level-k thinking. Subjects were further asked for some demographic information such as gender, field of study, and political attitude.

Each experimental session took around 60 minutes. The average payment was 19.79 Euro (minimum: 14.01 Euro, maximum: 32.90 Euro, median: 19.13 Euro).[18]

3.7 Results

In this section, I begin with a presentation of the tax decisions in the first round. In the second part, I consider the repeated decisions in the subsequent rounds. The third part deals with the effort provision and its evolution over time. Here, I identify the moral hazard of the higher redistributive tax. The comparison of Baseline and Tax_Random treatment identifies the selection effect. In the last two parts, I discuss different types of behavior, individual strategies, and decisions. Unless otherwise stated, I report bootstrapped standard errors, clustered at the session level in all regression results.

[17]To avoid framing effects, I did not use the terms "tax" and "redistribution" or any related wording in the experiment. Instead, I used "percentage" and "application of the percentage."

[18]This is above the minimum hourly wage of 12 Euro in Germany for June and July 2023.

3.7.1 Period-one behavior

The initial tax decision is made before knowing how (and by whom) each subject's individual income is generated for both the BASELINE and INCOME_RANDOM treatments. The initial tax decision is accordingly made behind a veil of ignorance, and I separate the analysis from the subsequent rounds, in which the income-generating process is well known.

The (null) Hypothesis 1 in section 3.3.3 states that subjects may choose $\bar{t}$ or $\underline{t}$. Testing this leads to my first result.

Result 3.1 *For both the BASELINE and INCOME_RANDOM treatments, subjects chose both tax rates in the first period.*

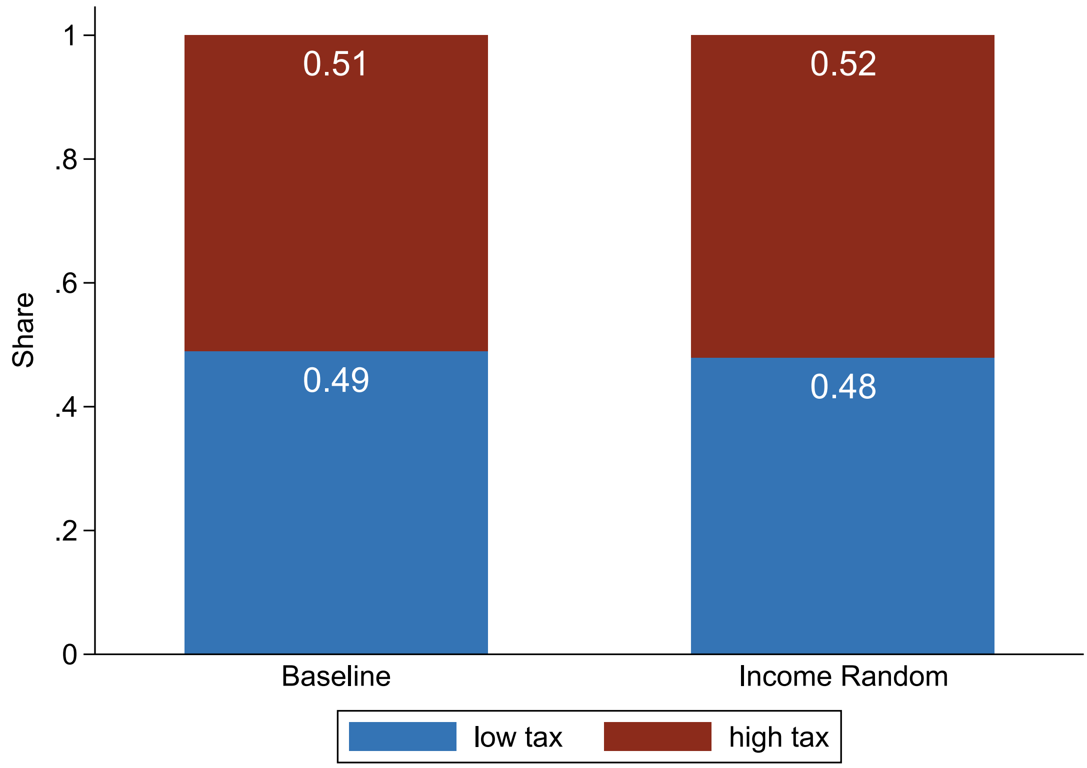


Figure 3.2: First-round decisions in BASELINE and INCOME_RANDOM.

In the experiment, subjects indeed chose both tax rates in the first period, as Figure 3.2 illustrates. In the first round of the BASELINE treatment, 51 percent of subjects chose $\bar{t}$ and accordingly 49 percent $\underline{t}$.[19] In the INCOME_RANDOM treatment with the same instructions before the first decision, the result is similar, with 52 percent choosing $\bar{t}$.[20] I find no significant difference in the first round decisions between the two treatments.[21] This 50:50 mixture of high and low redistribution found in the data is consistent with the results of other veil of ignorance studies. For example, Frohlich et al. (1987), Cappelen et al. (2007), and Deffains et al. (2016) also find very mixed results regarding the choice of different distributive rules that more or less insure against unfortunate (gross) income. The distributive rules used in these experiments, e.g., libertarian and egalitarian, can be interpreted as extreme cases of my low and high tax rates, which in contrast maintain some redistribution or some self-differentiating income, respectively.

3.7.2 Decisions over time

After the first round, subjects know the characteristics of the task and their own performance, receive information about their relative performance, and can update their initial beliefs accordingly. At this stage of the experiment, the insurance motive of choosing a high tax becomes less important as pointed out in subsection 3.3.4. Accordingly, starting with the second round, the choice of taxes should be driven by a different set of motives.

The hypotheses bundle 2a, 2b, and 2c relates to the tax decisions in experimental rounds two to six for the BASELINE treatment. The "immediate unraveling" Hypothesis 2a cannot be confirmed as Figure 3.3 illustrates. Hypothesis 2b states that voting by feet, which increases the proportion of low-tax choices, will occur over the course of the experiment. Hypothesis 2b is supported.

Result 3.2 *In the BASELINE treatment, subjects choose less redistribution as the experiment progresses.*

[19]Tax choices do not differ significantly from 0.5 (two-sided binomial test, p-value=0.906, based on the (initially independent) 96 observations of the first round), the share randomization implies.

[20]Tax choices do not differ significantly from 0.5 (two-sided binomial test, p-value=0.700, based on the (initially independent) 96 observations of the first round), the share randomization implies.

[21]Pearson's $\chi^2 = 0.0209$, df=1, p-value=0.885, based on the (initially independent) 192 observations of the first round.

The probability of choosing high tax decreases over the course of the experiment by 17.5 percentage points.[22] This result corresponds to a finding by Gerber et al. (2019) who note decreasing support for redistribution when subjects have more information about their own productivity.[23] This relates to the unraveling in my Baseline treatment which I formally test with Hypothesis 3 below.

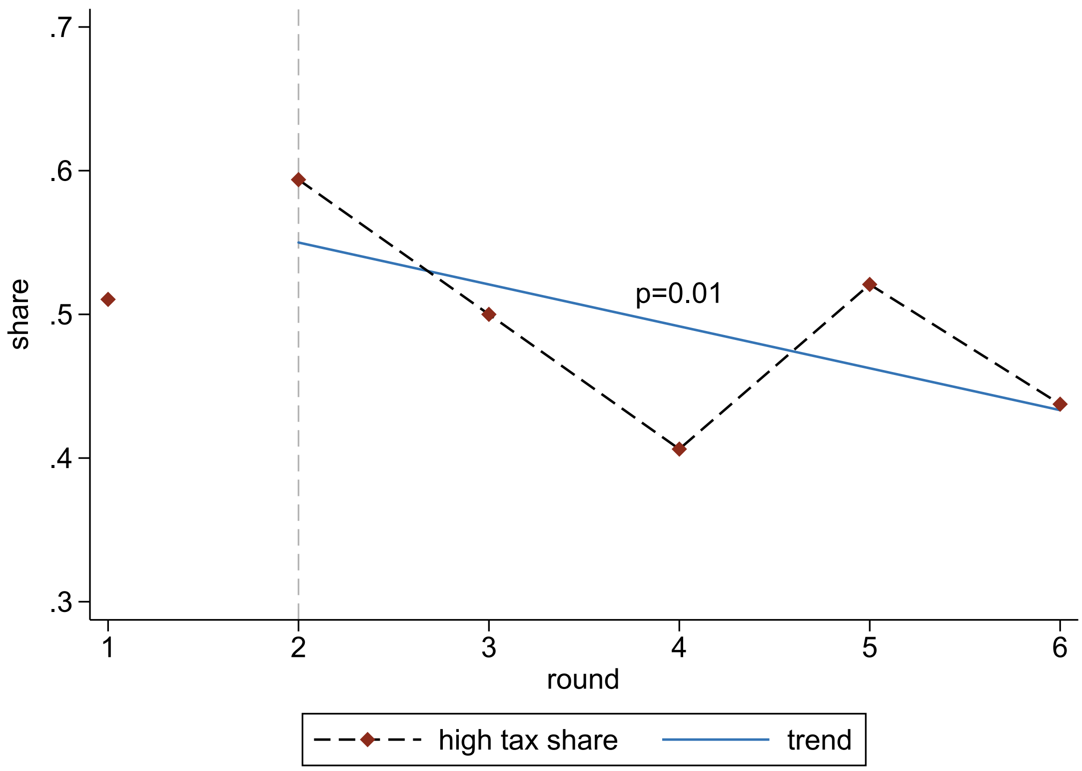


Figure 3.3: Relative frequency of the high tax rate over time (rounds) in the Baseline treatment.

I now test how the experience of redistribution, that is, whether a subject was a net payer or a net receiver (the subject benefited from the redistribution in the previous period) in the previous round, affects tax choice. This is formulated as Hypothesis 2c, which is also supported.

[22]This is tested using a linear probability model. The average marginal effect is $-.0292$ percentage points (*std. err.*=0.011) per round and significant (p-value=0.010). The result is reported in Table 3.2 column 3.

[23]They find that the probability of choosing the libertarian rule increases with more information, while the probability of choosing the proportional or egalitarian rule decreases with more information. See section 3.2.

Result 3.3 *In the* Baseline *treatment, net receivers (net payers) tend to choose the high tax rate* $\bar{t}$ *(the low tax rate* $\underline{t}$*) more often. This effect is more pronounced for those who previously chose the high tax rate.*

As Table 3.2 column 1 shows, subjects who were net receivers in the previous round are 18 percent and significantly more likely to choose the high tax (p-value=0.001). Obviously, it is the other way around for net payers. In the second column, I only consider subjects who previously chose the high tax, as I expect subjects not to switch back once they have chosen the low tax. I find that the probability of choosing high tax is 23 percent and significantly (p-value=0.012) larger for net receivers compared to net payers, a share even bigger than in the full sample. These results support the gradual unraveling and can be cautiously related to the results in Deffains et al. (2016), who find different redistribution preferences between "overachievers" and "underachievers" in a money-earning real-effort task. They find that subjects who earned above the median (referred to as "overachievers") preferred less redistribution, while others voted more for the egalitarian redistribution principle. Such self-serving behavior is also found by Durante et al. (2014).[24]

I have already shown that voting by feet favors the low tax rate as the proportion of subjects choosing the high tax rate decreases over time (Result 3.2). I now want to test whether this result can be related to the information unraveling effect in Baseline. Hypothesis 3 considers the time trends of subjects' tax choices in the Baseline treatment and the Income_Random treatment. In Table 3.2, column 4, I report the results for a time trend estimation of the two treatments Baseline and Income_Random.[25] Comparing tax choice time trends between these two treatments for rounds two to six, I find a significant negative time trend for tax choices in Baseline (*round*). The trend for the Income_Random treatment is not significantly different from that (*income_random* × *round*) which means that Hypothesis 3 is not supported. Although I do not find a difference in the time trends,

[24]Self-serving in the sense of not altruistically choosing more redistribution, but only when the subject benefits monetarily.

[25]Table 3.2 column 3 shows the time trend in the Baseline treatment which relates to (Result 3.2).

there is a significant level shift of 17.7 percentage-points for high tax rate choices in the INCOME_RANDOM treatment (*income_random*).[26]

Table 3.2: Tax choice (over time).

	high tax	high tax	high tax	high tax
net_receiver [$t-1$]	0.181***	0.227**		
	(0.0562)	(0.0901)		
round			-0.0292***	-0.0292**
			(0.0113)	(0.0122)
INCOME_RANDOM				0.177**
				(0.0827)
INCOME_RANDOM × round				-0.0229
				(0.0219)
Constant	0.415***	0.475***	0.608***	0.608***
	(0.0572)	(0.0704)	(0.0629)	(0.0696)
Treatment	BASELINE	BASELINE	BASELINE	BASELINE & INCOME_RANDOM
Condition		high tax in [$t-1$]		
Rounds	1-6	1-6	2-6	2-6
R^2	.0319233	.0515757	.0068074	.0216518
Observations	480	243	480	960

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *net_receiver*[$t-1$] is equal to 1 if the subject monetarily benefited from redistribution in the previous round. *round* reflects the time trend. INCOME_RANDOM is equal to 1 if the treatment is INCOME_RANDOM. INCOME_RANDOM × *round* is the interaction of INCOME_RANDOM and *round*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.7.3 Effort provision

For treatments and tax levels to matter, it is crucial that performance in the real-effort task is quite heterogeneous. This is indeed the case.[27] In the BASELINE treatment, the number of puzzles solved in one round ranges from 0 to 23 puzzles, with an average of 14.55 puzzles solved and a standard deviation of 2.86. For the TAX_RANDOM treatment, subjects' performances in the real-effort task vary between 2 and 23 with an average of 14.38 and a standard deviation of 2.69.

[26]A graphical representation can be found in Appendix 3.A.1.
[27]See Figure C.4 in Appendix 3.A.2.

In this section, I analyze the differences in effort (number of puzzles solved) when taxes are high vs. low for the BASELINE treatment and the TAX_RANDOM treatment. The difference in effort across tax regimes in the BASELINE treatment is driven by moral hazard and self-selection. There is no selection in the TAX_RANDOM treatment. Thus, the effort difference reflects the stronger moral hazard effect of a more redistributive tax. I then compare the differences between the two treatments, which allows me to disentangle the moral hazard effect and the selection effort in the BASELINE treatment.

3.7.3.1 Effort differences in Baseline

Subjects in BASELINE solved fewer puzzles at the high tax rate. This is consistent with Hypothesis 4. Figure 3.4 (left panel) plots the average number of puzzles solved (effort) by tax group. I find that the amount of effort provided in the low tax group is significantly higher than the effort provided in the high tax group.[28] This difference is 1.26 puzzles, which is the sum of the selection effect and the moral hazard effect of the higher tax.

Result 3.4 *In the BASELINE treatment, subjects who chose the high tax rate $\bar{t}$ solve fewer puzzles than those who chose the low tax rate $\underline{t}$.*

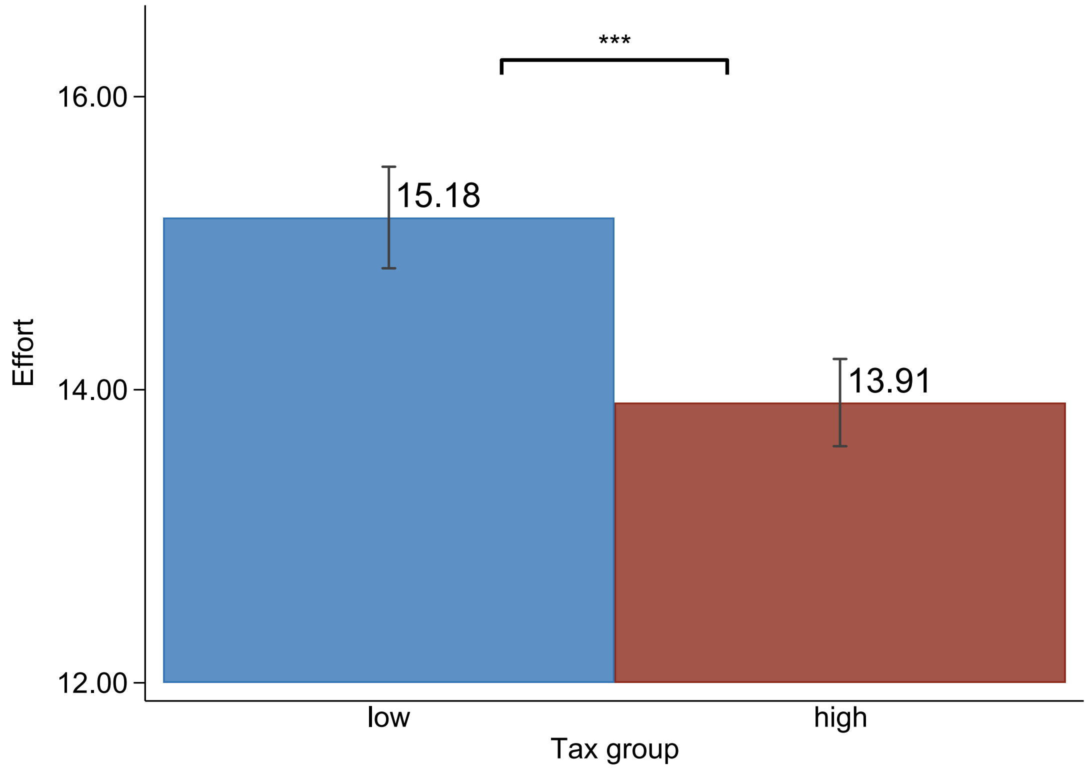
How does this difference emerge? The right panel of Figure 3.4 presents the average effort of these two groups over time. The two lines diverge over time. While the low tax group's performance increases due to self-selection and learning, the high tax group's performance is rather stable.[29] Given the general time trend of learning that is most prevalent in the early rounds,[30] this shows the high-effort subjects voting with their feet into the low-tax regime and, in contrast, the myopic self-selection of the low-effort subjects into the high-tax regime.[31]

[28] I use an OLS regression test for rounds 2-6, p-value=0.001.

[29] In the first round, when subjects chose the tax group behind a veil of ignorance, there is no significant difference in the effort provided between the tax groups (t-sample t-test with equal variances, p-value=0.901).

[30] See Appendix 3.A.5 for a more detailed analysis of learning effects.

[31] One solved puzzles more in the previous round decreases the probability of choosing the high tax rate on average by 4.8 percentage points (OLS, p-value<0.001). The selection effect is tested below with Hypothesis 5.



Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

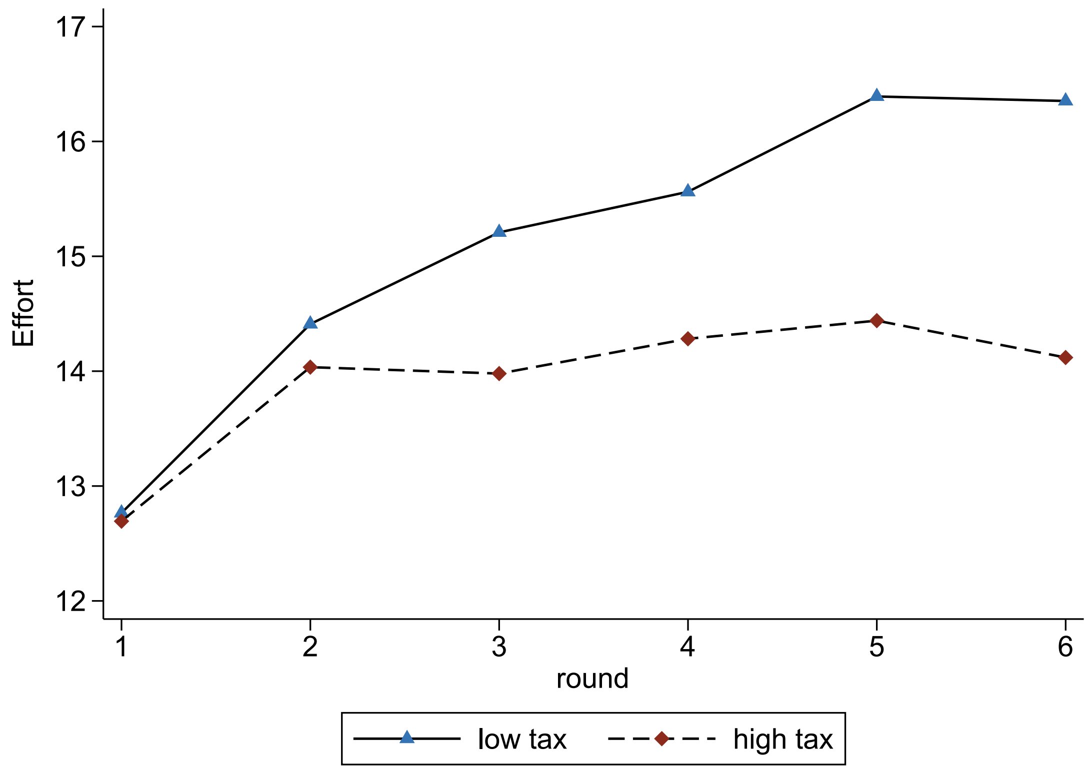


Figure 3.4: Average effort provided in each tax group for the Baseline treatment (upper part) and its evolution over time (lower part).

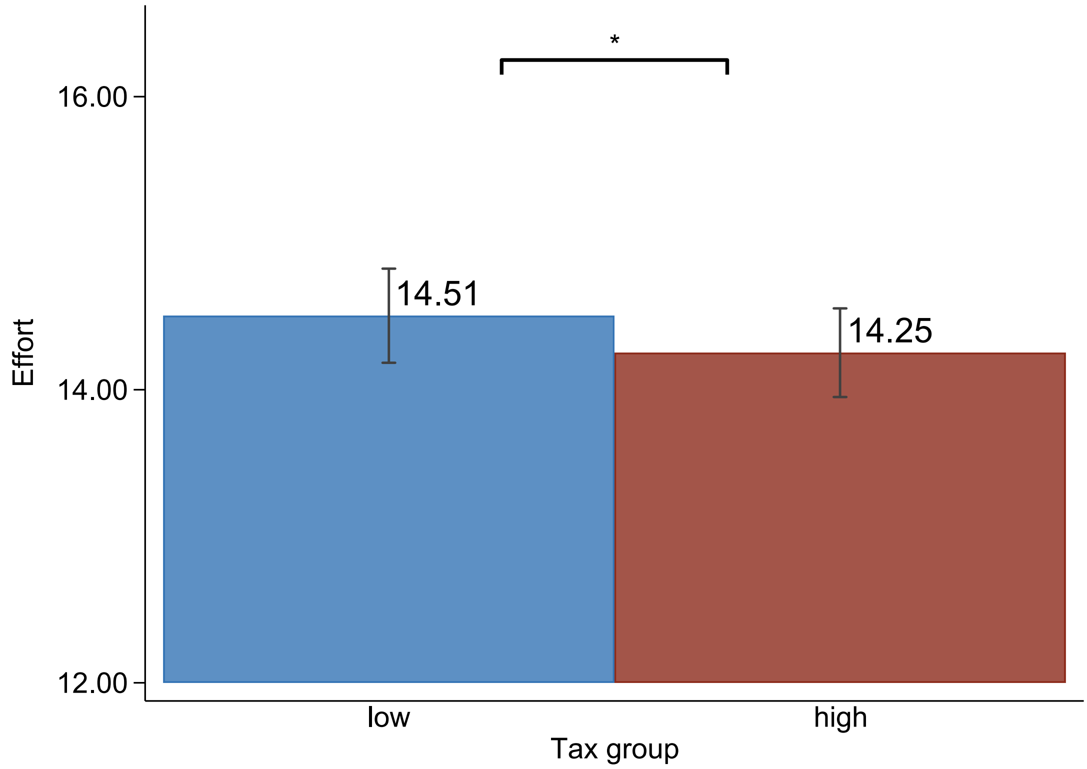
3.7.3.2 Effort differences in Tax_Random (moral hazard)

The difference in provided effort across the two tax regimes in the TAX_RANDOM treatment reflects a stronger moral hazard effect of the more redistributive tax. Figure 3.5 shows both the average effort differentiated by tax group and the effort provision over time for the two tax groups. I find a weakly significant difference between provided efforts in the two tax rate groups (left).[32]

The time trend is similar for both tax groups in the first half of the experiment, which is not surprising since subjects were randomly assigned to a group in each round.[33] For the second half, the spread is more pronounced between the two groups in Figure 3.5. This small difference between the two groups that slightly emerges in the second half of the experiment suggests that moral hazard effects might build up over time.

[32] I use an OLS regression testing for rounds 1-6, p-value=0.077 (for rounds 2-6, p-value=0.182).

[33] There is no significant difference for the trends of the two tax groups (p-value=0.306). This result is reported in Table 3.3, column 3.



Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

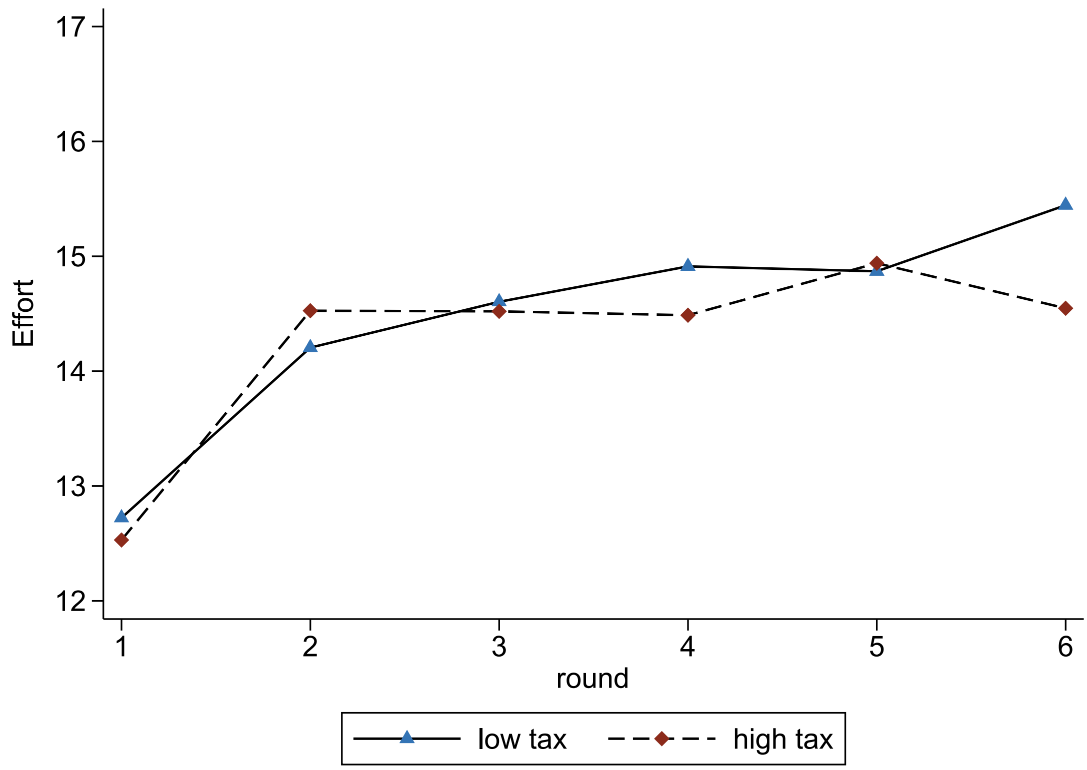


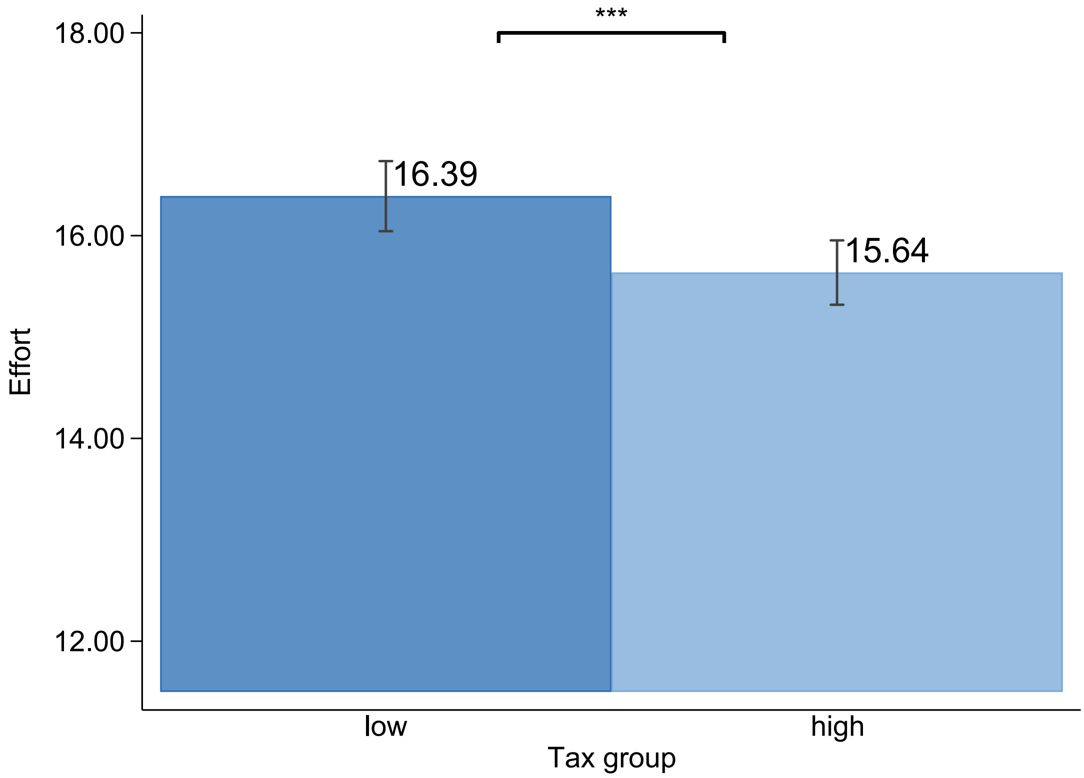
Figure 3.5: Average effort provided in each tax group for the TAX_RANDOM treatment (upper part) and its evolution over time (lower part).

Result 3.5 *In the* Tax_Random *treatment, there is a moral hazard effect for above-median performers, but not for below-median performers.*

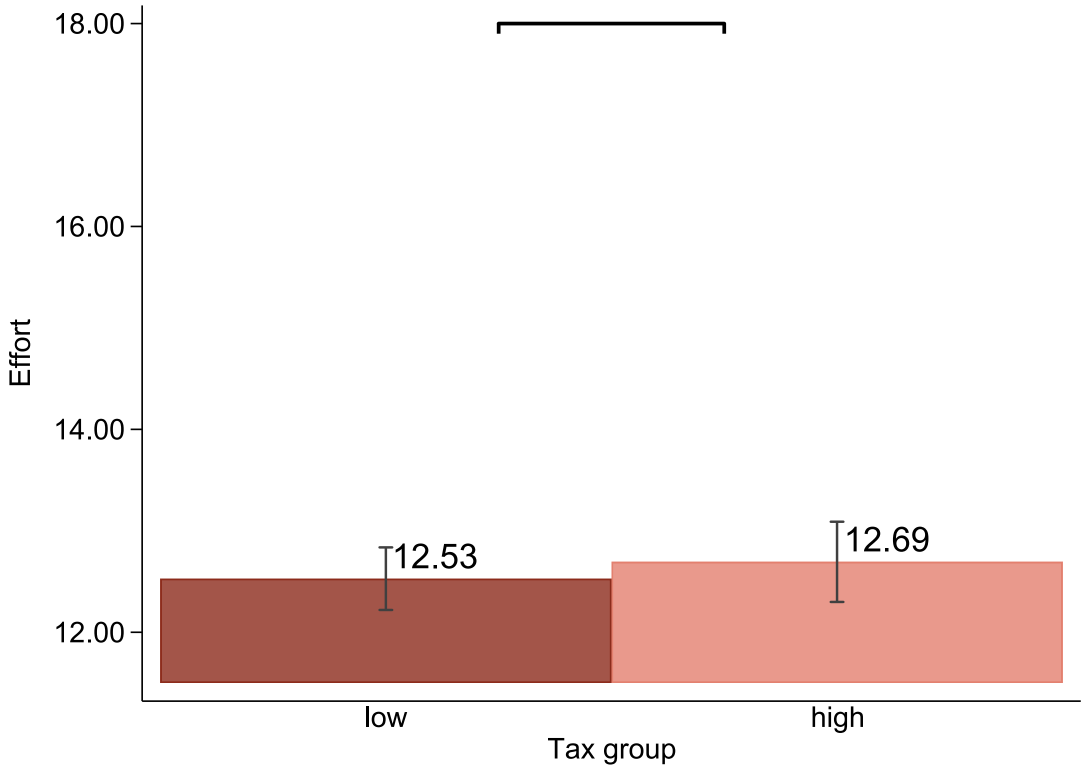
The moral hazard effect is largely driven by the high-effort subjects. To document this, I split the subjects in the Tax_Random treatment at the median of the number of puzzles solved into two groups and compare their effort in the two tax regimes.[34] This provides the moral hazard effect for an above-median performer group and a below-median performer group, which is presented in Figure 3.6. While I find no significant difference between the effort provision in the two tax regimes for the below-median group, this difference is significant for the above-median group.[35] The difference between these two estimated moral hazard effects is 0.92. This result suggests that the observed moral hazard effect is mainly driven by differences in the effort supply of above-median performing individuals between the two tax regimes. This shows the moral hazard difference between the two levels of redistribution.

[34] I perform the median split using the provided effort in the first round.

[35] I use an OLS regression test for rounds 1-6. For the below-median group, the estimated difference is 0.17 puzzles (*p*-value=0.581), and for the above-median group, the estimated difference is 0.75 puzzles (*p*-value=0.000).



Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.



Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 3.6: Average effort provided in each tax group for above-median performers (upper part) and for below-median performers (lower part). The median split was performed using the effort provided in the first round.

3.7.3.3 Selection effect

I now compare the effort provision over time between the two treatments BASELINE and TAX_RANDOM. Hypothesis 5 considers the differences in the effort differences of the tax groups between these two treatments. This is the selection effect. As shown graphically in the previous Figures 3.4 and 3.5, and the test results, the effort difference is larger and increases over time in the BASELINE treatment.

Result 3.6 *In the BASELINE treatment, there is a significant and substantial selection effect.*

Formally testing the differences, I find that the average effort gap between the groups is significantly larger (6.7 percent) in BASELINE compared to TAX_RANDOM.[36] This result is reported in Table 3.3, column (1).[37] The observed selection effect is substantial in magnitude and supports the voting by feet sorting into after unraveling.[38] The second column shows the same regression for the BASELINE treatment only, including the general time trend to control for e.g., learning effects. In column 3, I report the regression results for TAX_RANDOM.

The coefficient *high_tax* indicates the sum of the (negative) selection effect and moral hazard effect for the high tax rate as it captures the difference to the low tax group in the BASELINE treatment. The sum of *high_tax* and *high_tax* × TAX_RANDOM reflects the size of the average moral hazard effect (0.253). The sign matches the prediction of less effort when taxes are higher.[39]

[36]This is the difference between the effort difference in BASELINE and the effort difference in TAX_RANDOM without controlling for time effects (1.01 puzzles).

[37]This test was performed using rounds one through six as in the previous results of this subsection. If I exclude the first round, the coefficient is slightly larger at 1.24 puzzles and also significant (p-value=0.013).

[38]Table C.1 in Appendix 3.A.2 presents the relationship between the number of puzzles solved in the previous round and the tax selection in the current round. The findings reveal a negative correlation between a greater number of previously solved puzzles and the selection of the high tax rate group in the BASELINE treatment. This shows how the unraveling of the previous task performance influences the next tax decision and thus the voting by feet decision.

[39]I use an OLS regression testing for rounds 1-6, p-value=0.086 (for rounds 2-6, p-value=0.204). Participants' post-experimental feedback anecdotally supported the moral hazard observation, as they commented, for example, "The experiment was very tiring and the motivation to solve these puzzles for 6 rounds 3min dwindled", or "For the last round I didn't bother to solve the Crypto task, since 85% were mostly losing compared to 15%."

Table 3.3: Comparison of effort provision in Baseline and Tax_Random.

	effort	effort	effort
high_tax	-1.263***	0.378	0.309
	(0.427)	(1.008)	(0.302)
Tax_Random	-0.670		
	(0.670)		
high_tax	1.010**		
× Tax_Random	(0.459)		
round		0.687***	0.459***
		(0.148)	(0.0371)
high_tax		-0.439*	-0.135
× round		(0.226)	(0.0978)
Constant	15.18***	12.69***	12.85***
	(0.566)	(0.835)	(0.317)
Treatment	Baseline & Tax_Random	Baseline	Tax_Random
Rounds	1-6	1-6	1-6
R^2	.0278598	.1447726	.0658196
Observations	1152	576	576

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *high_tax* is equal to 1 if the high tax rate was chosen. Tax_Random is equal to 1 for the Tax_Random treatment. *high_tax* × Tax_Random is the interaction of *high_tax* and Tax_Random. *round* reflects the time trend. *high_tax* × *round* is the interaction of *high_tax* and *round*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

For both treatments, I do not find a significant effect of group size on individually provided effort levels. This finding does not support Hypothesis 6 which assumed a negative impact of group size on individual effort provision.[40]

3.7.4 Behavioral analysis

I now turn to a more behavioral analysis in order to identify potential factors driving or counteracting the observed effects. The individual choice is influenced by several overlaying (behavioral) factors. As Figure 3.1 in section 3.3.3 shows, individual risk preferences influence the initial tax decision. In the experimental setting the veil of ignorance is not completely sealed as subjects hold beliefs based on self-experience. Even if the income-generating process or task is not known, subjects bring their own history and experiences into the lab and can be, for example, over- or underconfident regarding their own abilities. Subjects are heterogeneous agents who bring their own talents and characteristics to the lab.

In addition to overconfidence and underconfidence in the first round, I also elicited several behavioral measures after the main experiment. I elicited a more general measure of confidence based on quizzes and performance beliefs. Using a beauty contest, I also elicited a measure of level-k thinking. I further added a dictator decision to elicit a measure of altruism. These elicitations, as well as the measure of risk, which all occurred after the main experiment, were incentivized. I then added questions about subjects' attitudes toward talent and fairness, their self-reports of altruism, and their political views. In the last part, I collected demographics such as field of study, gender, age, and experience with economic experiments as well.

In this subsection, I analyze the effect of various behavioral factors on tax choices for both the first and subsequent rounds. I also relate behavioral factors to the provision of effort.

[40]This is tested using a linear probability model. For the Baseline treatment the p-value of testing the group size effect on effort is p-value=0.655 and for Tax_Random p-value=0.363.

3.7.4.1 First round

Hypothesis 7 relates individual confidence and risk preferences with initial tax choices. To this end, I run linear probability regressions, clustered at the session level. I define an overconfident and underconfident subject .[41] For risk attitudes, I use the test of risk aversion as in Charness and Villeval (2009); Charness and Gneezy (2010) to elicit risk preferences. A higher variable risk score reflects less risk aversion. The highest possible risk value reflects a risk-neutral attitude.

Result 3.7 *In the* Baseline *treatment, underconfident subjects are more likely to choose the high tax rate.*

I find some limited support for the expected effect of confidence on the initial tax decision, but no support for the part on risk attitudes. Table 3.4 shows subjects' initial tax choice when I control for risk preferences and overconfidence and underconfidence.[42] Combining the first round decisions of Baseline and Income_Random,[43] I find that underconfident subjects are more likely to choose the high tax. This result is significant (p-value=0.074). I find no significant effect for overconfident subjects (p-value=0.208). Risk preferences also have no significant effect on the probability of choosing the high tax (p-value=0.700). These results are shown in Table 3.4, column 1. In column 2, I include a gender dummy reflecting that male subjects are significantly less likely than female subjects to choose the high tax rate in the initial decision (p-value=0.080).[44] Separately estimating the two treatments, I find mixed

[41]These measures result from a belief elicitation after the first task was completed. Before providing the first feedback, I asked subjects whether they believed they performed better, the same, or worse than the average in their tax group. Subjects who reported "better than average" but performed at or below average are labeled as overconfident, while subjects who believed they were "worse than average" but actually performed at or above average are labeled as underconfident.

[42]These confidence measures are elicited via a questionnaire after the first round of the task, which asks whether subjects believe they are better, equal, or worse than average. I assessed these self-reports against actual relative performance, and defined the dummy variables to take the value of one if self-report and actual performance were misaligned, such that the better-than-average statement was associated with below-average performance in the task, and vice versa.

[43]This first decision was made under same conditions, which means before learning about the income-generating process which is, of course very different).

[44]In this regression, the estimated effect for $underconfident$ subjects is also significant (p-value=0.053). Estimating the effect separately for male and female subjects, I find that among male subjects, those classified as overconfident are less likely to choose the high tax rate in the initial decision (this must be interpreted with caution, as the number of overconfident subjects is

results for risk preferences' effects, while estimated effects resulting from the over- and underconfidence of subjects show the intuitively expected directions.[45]

Table 3.4: Risk and confidence.

	high tax	high tax	high tax	high tax
risk	0.000470	0.000597	0.00301***	-0.00235
	(0.00122)	(0.00129)	(0.00104)	(0.00166)
overconfident	-0.250	-0.243	-0.208	-0.524***
	(0.199)	(0.170)	(0.269)	(0.0686)
underconfident	0.194*	0.199*	0.116	0.282**
	(0.109)	(0.103)	(0.170)	(0.143)
gender		-0.0944*		
		(0.0539)		
Constant	0.468***	0.499***	0.316***	0.641***
	(0.0931)	(0.0941)	(0.0600)	(0.148)
Treatment	BASELINE & INCOME_RANDOM	BASELINE & INCOME_RANDOM	BASELINE	INCOME_RANDOM
Rounds	1	1	1	1
R^2	.0234207	.0310924	.0449281	.0592448
Observations	192	190	96	96

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *risk* takes values between 0 and 100 with a step size of 0.5. *risk* is equal to 0 for full risk aversion and is equal to 100 for risk neutrality. *overconfident* is equal to 1 if a subject stated that her performance was above average while it actually was below or equal to average performance. *underconfident* is equal to 1 if a subject stated that her performance was below average while it actually was above or equal to average performance. *overconfident* and *underconfident* are elicited after the real effort task in the first round of play. *gender* is equal to 1 if a subject has declared itself to be male. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.7.4.2 Subsequent rounds

In addition to risk and confidence, which are expected to be the most prevalent effects for tax choice behind the veil of ignorance, other behavioral factors may also influence individual decisions in the experiment over time. As uncertainty is

small). Among female subjects, I find that those classified as underconfident are significantly more likely to choose the high tax rate (p-value=0.012). This is consistent with the gender confidence gap findings in the literature (see Möbius et al., 2022, and the references therein).

[45] These results are shown in Table 3.4, columns 2 and 3. I also analyzed the effects of other behavioral factors which I elicited on the tax decision in the first round. These results can be found in Appendix 3.A.3.1.

reduced by the unraveling after the first round, individual (behavioral) characteristics may reveal significant differences between subjects. I test for the effect of various characteristics, such as altruism, level-k thinking, as well as attitudes toward talent and fairness, political views, field of study, age, experience, or gender. The results are reported in Table C.7 in Appendix 3.A.3. I control for the time trend, as this has a striking impact on tax choices and also effort.

For most of the behavioral measures, I find no significant effect. One exception is $underconfident$. This variable shows the strongest and most significant effect on tax choice in the Baseline and Income_Random treatments. This shows that underconfident subjects (as classified based on the first round) are still more likely to choose the high tax rate in subsequent rounds. At the same time, they do not put in less effort than the others in the Baseline treatment. I also find a significant effect of experience as an experiment participant on the high tax choice in the Baseline treatment. Subjects with more experience are less likely to choose the high tax rate compared to less experienced subjects. In terms of effort, however, there is no significant difference between more and less experienced subjects in this treatment. More can be found in Appendix 3.A.3.2.

3.7.5 Individual-level analysis

Concluding, I take a look at the tax-choice data at the individual level. Are there specific types of players? Are there persistent patterns?

Each participant can be represented by a decision vector $v^i = (v^i_1\, v^i_2\, v^i_3\, v^i_4\, v^i_5\, v^i_6)$ where $v^i_r \in \{high, low\}$ is participant i's tax choice in round r. In total, there are potentially $2^6 = 64$ different decision vectors. In the data, I find 45 different vectors in Baseline and 38 in Income_Random.[46] So, participants choose their tax level rather differently. In the following, I consider data from Baseline unless otherwise stated.

As the data are quite heterogeneous, I look for decisions vectors that occur

[46]There were 96 individual subjects in each treatment.

Table 3.5: Strategies in BASELINE.

Strategy	BASELINE
alternator	1
always low	6
always high	7
start low switch to high	8
start high switch to low	12
start and keep low except try high once	9
start and keep high except try low once	3
start and keep low except try high twice (separately)	8
start and keep high except try low twice (separately)	5
start and keep low except try high for 2 consecutive rounds	3
start and keep high except try low for 2 consecutive rounds	7

Notes: This table shows the number of subjects who applied different specified strategies in BASELINE.

frequently. The most frequent decisions vector can be labeled "always high" and occurs seven times, meaning that seven subjects chose the high tax rate in every round. The second most frequent decisions vector reads "always low" (six occurrences). Vectors with other patterns occur less frequently, with 22 of the 45 vectors observed occurring only once.

The next step would be to identify strategies that comprise different decisions vectors with qualitatively similar patterns or sequences. A prominent strategy reads "choose the high tax rate in round one but eventually switch to low," which, like the "always low" strategy, fits the findings presented in Result 3.3. This strategy includes the five decisions vectors that begin with high and then (in any round two to six) switch to low. With 12 occurrences, this is the most frequent strategy. The opposite strategy, start low and switch high, occurs only eight times.

Another strategy that comes to mind is the "alternating" strategy. The alternating strategy between high and low occurs only once, while the alternating strategy in the opposite direction cannot be observed in BASELINE. Nevertheless, this strategy should not be popular for this treatment because of the unraveling and informed self-selection.

As Hypothesis 2c suggests, individual tax decisions can depend on whether a subject was a net payer or net receiver in the previous rounds. The strategy which can be called "if net payer in previous round, choose low tax; if net receiver in previous round, choose high tax" was adopted seven times, three of these played "always high" or "always low." The opposite strategy was adopted three times, two of these played "always high."[47] In total 36 subjects followed this strategy, when I allow for one deviation.

Over the course of the six rounds, the majority of subjects experienced themselves as both net payers and net receivers. However, out of a total of 96 subjects in Baseline, 22 were consistently net payers in every round. Conversely, 15 subjects were net receivers in all rounds. These net payers chose the low tax rate more often on average.[48] Similarly, net receivers chose the high tax rate more often.[49]

3.8 Conclusion

Mounting evidence suggests that taxes can affect the location choices of people, both within and across countries. This channel of migration can lead to further tax efficiency costs that policymakers need to consider when setting tax policy (Kleven et al., 2020). In particular, high-income individuals are likely to respond to relatively higher taxation by moving to other jurisdictions (see, e.g., Schmidheiny and Slotwinski, 2018; Martínez, 2022).

From the perspective of the individual, the question is: which side to choose? The decision to locate on one side or the other of a jurisdictional border with different income tax rates has long been a consideration for individuals. In today's era of increased globalization, increased migration, and the growing prevalence of remote work, this choice has taken on even greater significance. At the individual level, this

[47] For subjects choosing "always high" there is no noticeable pattern, when I count the deviations from the "if net payer in previous round, choose low tax; if net receiver in previous round, choose high tax" strategy except that two subjects had been net payers and choose high tax all the time. This could be interpreted as a clear preference for the high redistribution of these subjects.

[48] 3 subjects always chose the low tax rate from the second round on. When I allow for one deviation, this is true for 10 subjects.

[49] 2 subjects always chose the high tax rate from the second round on. When I allow one deviation, this is true for 8 subjects.

question can be answered by voting by feet.

In this laboratory experiment, participants are first asked to choose between two tax systems (high or low) that focus solely on redistribution, and then they engage in an income-generating task that requires real effort. These tax systems have contrasting effects: high taxes provide insurance against unfavorable gross incomes, but also create the risk of reduced effort due to moral hazard. A laboratory experiment is a suitable method to address my research questions. The controlled setting allows for the observation of both, moral hazard and selection, isolated from other intervening factors in a real-life setting.

Answering the questions of what tax rate subjects will choose and how tax choices change over time, I distinguish between initial and subsequent tax choices. The results of my experiment show that participants' initial tax choices, made without knowledge of their personal circumstances, vary significantly, as the proportion of high and low tax choices is nearly 50:50. Those who exhibit underconfidence are more likely to choose higher tax rates. As participants become familiar with the task and their relative income levels in subsequent rounds, they adjust their tax preferences opportunistically. High-income individuals tend to favor the lower tax rate, while low-income participants initially tend to favor the higher tax rate. This shifts toward lower taxes over time and can be described as a form of voting by feet.

I also find that net recipients (payers) more often choose the high (low) tax rate in the next round. This behavior is consistent with the "gradual unraveling" hypothesis (H2b). Net receivers naively (or myopically) choose the high tax rate because they expect to continue receiving in the future. But the selection effect (high performers choosing the low tax) eventually turns the former net receivers into net payers. At that point, they switch to the low tax rate.

I address the endogeneity of tax choice and effort over time in my Baseline treatment using two treatments where one of the two variables is exogenously given. With this design, on the one hand, I examine whether unraveling affects tax choice in my baseline treatment relative to the treatment in which income is uncertain as randomly assigned. The phenomenon of "unraveling" becomes evident after the first round, as subjects start to become aware of their relative performance over time. On the other hand, I control for the selection effect in my baseline treatment relative to a treatment in which tax choice is exogenously given. This control group helps to illustrate the effects of selection and moral hazard in the context of the Baseline treatment. Testing for differences in moral hazard between high- and low-effort subjects in my Tax_Random treatment to identify potential drivers, I find that the moral hazard is driven solely by high-effort subjects adjustments to their effort provision given different redistribution levels.

An analysis of individual behavior reveals different strategies used by participants in choosing between the two tax rates. While the strategy of always choosing the same tax rate is also prominent, the strategy of starting with a high tax and switching to a low tax is the most prominent. Allowing for one deviation, 36 out of 96 subjects followed the premise of choosing high tax if they were previously net receivers and low tax if they were previously net payers when choosing the tax rate.

Overall, the results of this study show that in a repeated setting, individuals tend to choose less redistribution over time. By combining voting by feet with a real-effort task, this experiment provides insight into the evolving preferences for redistribution over time. As individuals become more aware of their actual and relative performance, redistributive choices decrease due to unraveling and self-selection. These findings provide valuable insights not only for the academic debate on redistributive preferences, but also for real-world issues.

Translating my finding that people opportunistically choose redistribution over time into a real-world context, I provide experimental evidence on issues such as location choice at the border of different income tax jurisdictions, or tax evasion by high-income individuals. Especially in an era when highly skilled individuals have increased flexibility in their location choices due to remote work options, this "voting by feet" may increase.

Appendices

Appendices

3.A Supplementary results

3.A.1 Decisions over time

The average tax rate choices in each round for the Baseline treatment and the Income_Random treatment are shown in Figure C.1. This shows the relatively high level of high tax choices for the Income_Random treatment in the early rounds of the experiment. The included confidence intervals for both treatments express the high variation across sessions. Figures C.2 and C.3 show the share of high tax decisions over time for each session for the two treatments Baseline (Figure C.2) and Income_Random (Figure C.3).

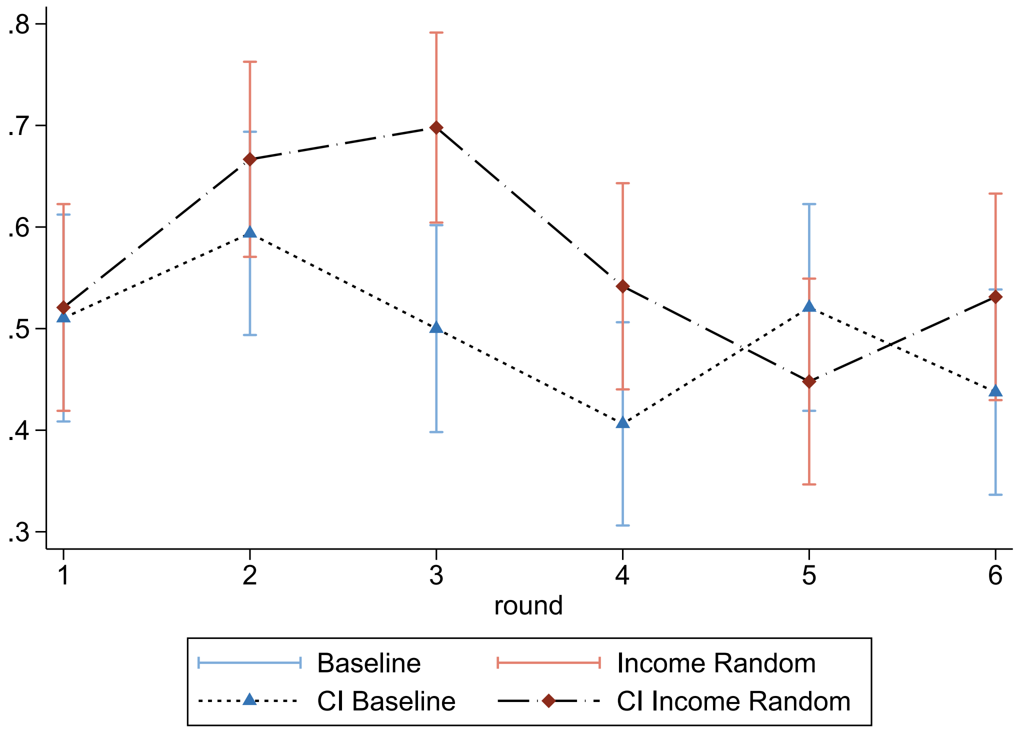


Figure C.1: Average high tax rate choices over time (rounds) for the Baseline treatment and the Income_Random treatment.

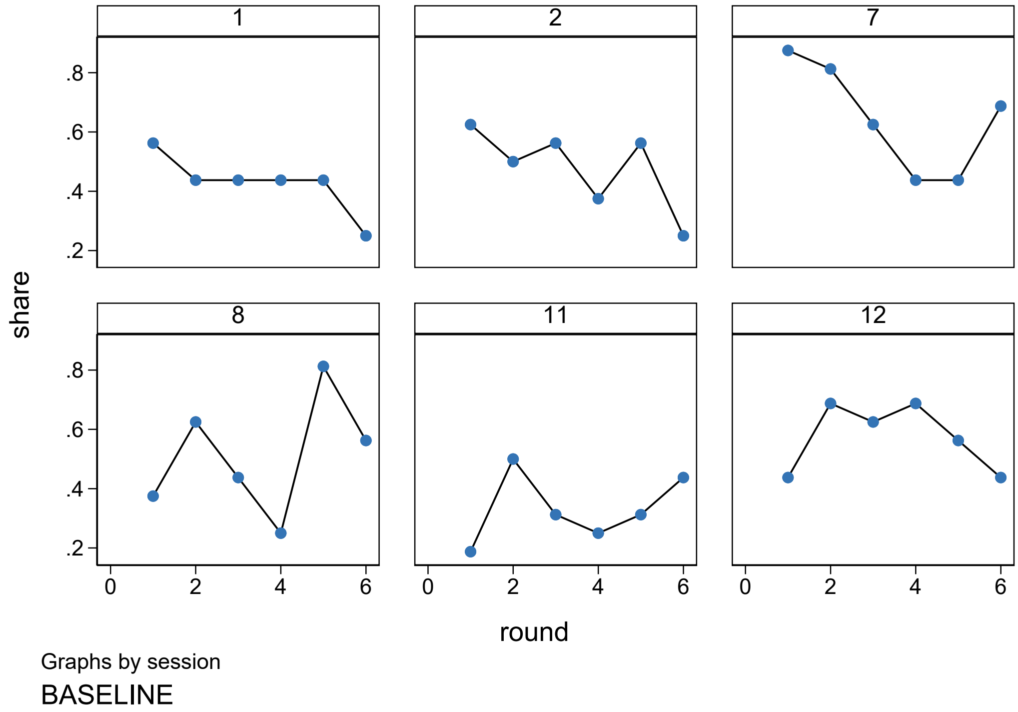


Figure C.2: Relative frequency of the high tax rate over time (rounds) by session for BASELINE treatment.

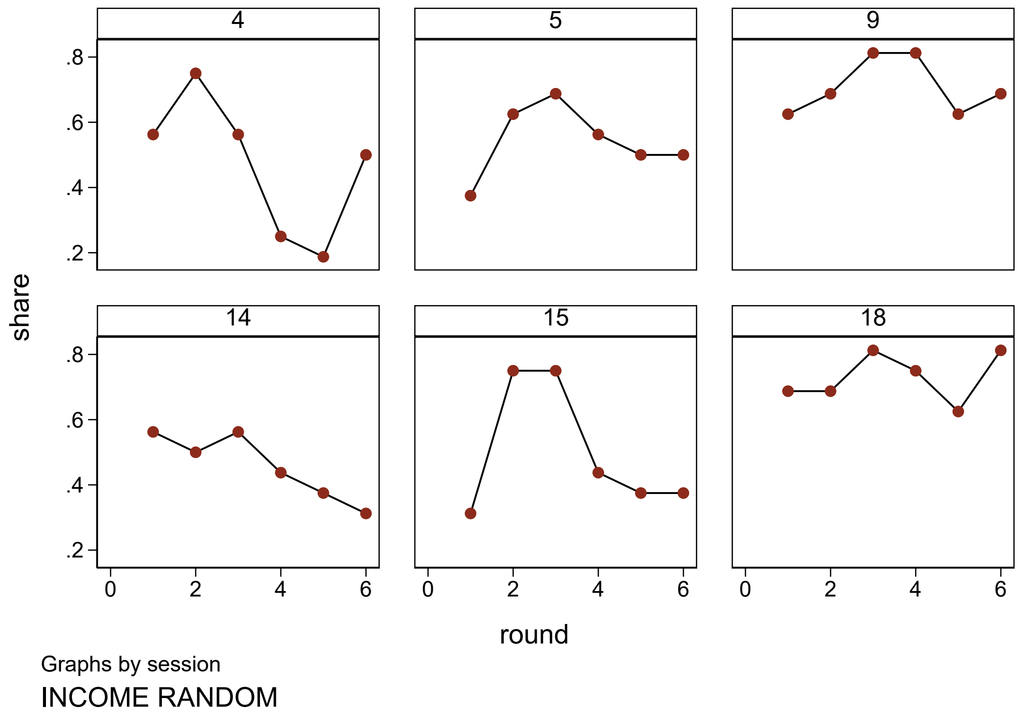


Figure C.3: Relative frequency of the high tax rate over time (rounds) by session for INCOME_RANDOM treatment.

3.A.2 Effort provision

Effort provision is heterogeneous within sessions and treatments. For BASELINE and TAX_RANDOM, Figure C.4 shows the mean effort of each participant, including confidence intervals.

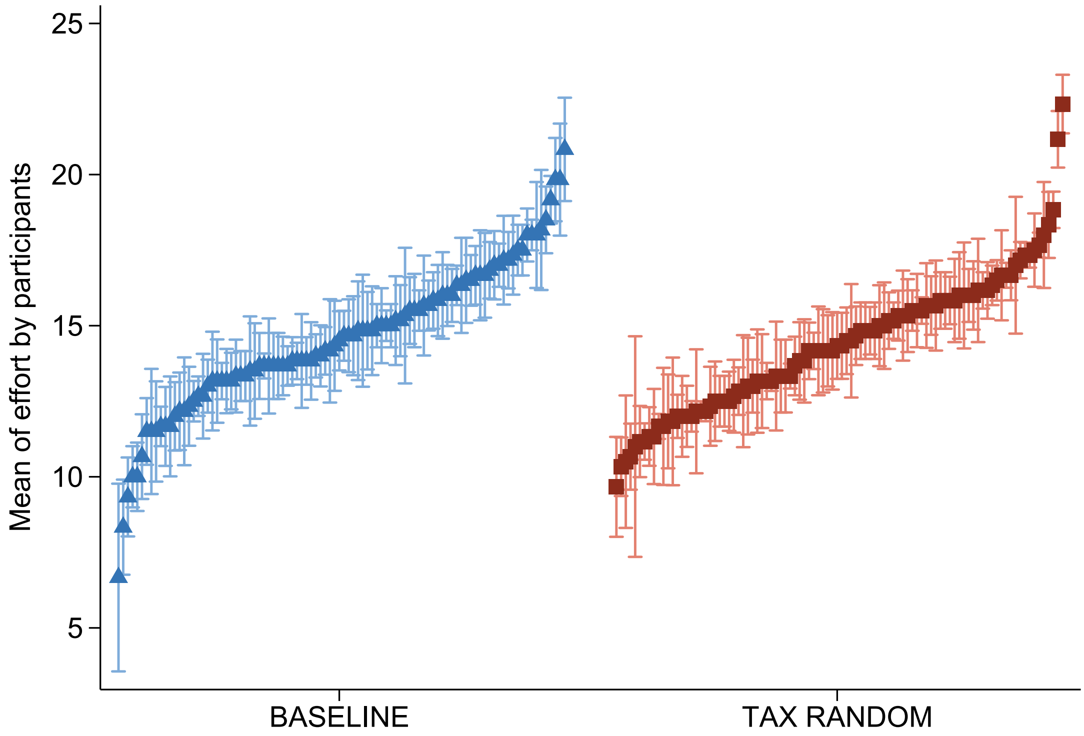


Figure C.4: Mean effort of each participant in the BASELINE treatment (left) and TAX_RANDOM treatment (right).

Table C.1 presents the relationship between subjects' high tax rate choices and the number of puzzles they solved in the previous period. The first column shows that a higher number of previously solved puzzles significantly lowers the likelihood of choosing the high tax rate in the BASELINE treatment. The same relationship cannot be confirmed for the TAX_RANDOM treatment. The probability that a subject chooses the high tax rate decreases with each additional previously solved puzzle by 4.5 percentage points more in the BASELINE treatment compared to the TAX_RANDOM treatment.

Table C.1: Previous effort and tax choice.

	high_tax	high_tax	high_tax
Effort [$t-1$]	-0.0480***	-0.00148	-0.00148
	(0.0116)	(0.0108)	(0.0113)
BASELINE			0.669***
			(0.238)
Effort [$t-1$] × BASELINE			-0.0465***
			(0.0171)
Constant	1.181***	0.513***	0.513***
	(0.145)	(0.162)	(0.172)
Treatment	BASELINE	TAX_RANDOM	BASELINE & TAX_RANDOM
Rounds	1-6	1-6	1-6
R^2	.0751294	.0000617	.0375955
Observations	480	480	960

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. $Effort$ [$t-1$] is the effort provision (number of correctly solved puzzles) in the previous period. BASELINE is equal to 1 for the BASELINE treatment. $Effort$ [$t-1$] × BASELINE is the interaction of $Effort$ [$t-1$] and BASELINE. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.A.3 Behavioral analysis

3.A.3.1 First round

I tested for the effect of several behavioral effects in the first round. When tested individually, no effect is striking. I find significant effects for risk (*risk*) and altruism (*keep* and *altruism_wp*13459). However, these are small and of negligible magnitude. When I join all potential behavioral explanations in one regression, I again find very small and mainly not significant effects in the BASELINE treatment. What stands out, is the effect of being *overconfident* in the INCOME_RANDOM treatment.[50] This suggests that an overconfident attitude toward own performance relative to the group average may have lead to more low tax choices. This might be an overestimation of individual luck.

[50]Although it is for (their own) income totally irrelevant, subjects also performed the real-effort task in the INCOME_RANDOM treatment.

Table C.2: Behavioral effects in round one (single) I.

	high tax	high tax	high tax	high tax	high tax	high tax	high tax	high tax
risk	0.00300***							
	(0.000989)							
guess		0.00543*						
		(0.00284)						
keep			-0.000252***					
			(0.0000891)					
altruism_wp13421				0.0302				
				(0.0281)				
altruism_wp13459					0.000466***			
					(0.000146)			
overconfident						-0.183		
						(0.238)		
underconfident							0.142	
							(0.161)	
confidence								0.0253
								(0.0233)
Constant	0.323***	0.278	0.700***	0.300	0.428***	0.516***	0.494***	0.472***
	(0.0764)	(0.177)	(0.104)	(0.268)	(0.107)	(0.0896)	(0.0887)	(0.107)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Rounds	1	1	1	1	1	1	1	1
R^2	.0335406	.0651381	.0185768	.0195282	.0439576	.004048	.0082148	.0062951
Observations	96	96	96	96	96	96	96	96

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. More detailed information on the variables can be found in Table C.8.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Behavioral effects in round one (single) II.

	high tax	high tax	high tax	high tax	high tax	high tax	high tax	high tax
politics	-0.00339							
	(0.0165)							
talent_fair		0.0130						
		(0.0136)						
not_only_talent			0.0229					
			(0.0308)					
not_only_talent2				0.00501				
				(0.0108)				
age					0.000936			
					(0.00694)			
gender						-0.0611		
						(0.0588)		
business/economics							-0.173	
							(0.115)	
num_experiments								-0.0504
								(0.0690)
Constant	0.521***	0.441***	0.332*	0.482***	0.482**	0.594***	0.557***	0.638***
	(0.0581)	(0.107)	(0.181)	(0.0680)	(0.209)	(0.110)	(0.0800)	(0.177)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Rounds	1	1	1	1	1	1	1	1
R^2	.0001735	.0049665	.0073826	.0008163	.0001766	.0035027	.0235231	.0105618
Observations	96	96	96	96	93	96	96	96

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. More detailed information on the variables can be found in Table C.8.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Behavioral effects in round one (combined).

	high tax	high tax	effort	effort
risk	0.00339***	-0.00132	-0.0133	-0.00669
	(0.00128)	(0.00220)	(0.0105)	(0.00986)
guess	0.00478	-0.00296	-0.00872	-0.00820
	(0.00292)	(0.00277)	(0.0108)	(0.0174)
keep	-0.00000834	-0.0000667	-0.000194	0.0000983
	(0.000193)	(0.000107)	(0.000519)	(0.00135)
altruism_wp13421	-0.00414	0.00385	-0.0997	0.0403
	(0.0319)	(0.0319)	(0.109)	(0.164)
altruism_wp13459	0.000580***	0.0000983	-0.000918	-0.00229
	(0.000222)	(0.000259)	(0.000799)	(0.00195)
overconfident	-0.313	-0.698***	-1.385	0.373
	(0.284)	(0.218)	(0.853)	(1.554)
underconfident	0.156	0.217	0.664	2.376***
	(0.196)	(0.169)	(0.891)	(0.666)
confidence	-0.00739	-0.0220	-0.242	0.00919
	(0.0229)	(0.0206)	(0.179)	(0.199)
politics	-0.00818	-0.0831*	0.0996	0.239
	(0.0272)	(0.0442)	(0.116)	(0.286)
talent_fair	0.0226**	0.0370	-0.0236	0.164*
	(0.00987)	(0.0248)	(0.0586)	(0.0912)
not_only_talent	0.0314	-0.0239	0.0643	-0.0541
	(0.0456)	(0.0313)	(0.152)	(0.190)
not_only_talent2	-0.0171	-0.0243	-0.0870	-0.174
	(0.0190)	(0.0188)	(0.144)	(0.149)
age	0.00131	0.0130	-0.144***	-0.0843
	(0.00747)	(0.00919)	(0.0556)	(0.0532)
gender	-0.0203	-0.131	1.261	-0.339
	(0.117)	(0.104)	(0.844)	(0.611)
business/economics	-0.123	0.0389	-1.238*	1.015
	(0.136)	(0.126)	(0.722)	(0.884)
num_experiments	-0.0381	-0.125***	0.562	1.351***
	(0.0808)	(0.0373)	(0.514)	(0.333)
Constant	-0.0922	1.313***	15.86***	11.43***
	(0.691)	(0.436)	(2.392)	(3.006)
Treatment	BASELINE	INCOME_RANDOM	BASELINE	TAX_RANDOM
Rounds	1	1	1	1
R^2	.205491	.2619569	.3240427	.4277901
Observations	93	77	93	79

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. More detailed information on the variables can be found in Table C.8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.A.3.2 Subsequent rounds

The following two tables, C.5 and C.6, show the regression results for the variables used in Table C.7 when solely used as independent variables to explain the share of high tax choices. In this stand-alone analysis, the variables *num_experiments*, which indicates the experience as an experiment participant, and $underconfident$, which is a measure taken after the first real-effort task when subjects were asked about their relative performance belief compared to the group's average, provide the strongest explanatory power according to the reported R^2.

Table C.7 reports the results for a joint regression of all behavioral or demographic variables. In the joint regression for all factors, I find no significant or null effects on the tax choice for risk preferences, level-k thinking or altruism, in neither the form of a dictator game and self-statements. This is also the case for effort as a dependent variable.[51] Concerning confidence measures, I find significant effects for the over- and underconfidence measure elicited after the first real-effort task. More underconfident subjects are significantly more likely to choose the high tax rate in the Baseline treatment (column 1). This is also the case for the Income_Random treatment (column 2). This can be well linked to the previous finding in subsection 3.7.4.1 and shows that the effect of underconfidence in the task persists. These over- and underconfidence measures elicited after the first round are also significant concerning the effort provision in subsequent rounds in the Tax_Random treatment (column 4). For the confidence measure that I elicited using quizzes and performance beliefs after the main experiment, I find no significant effects.

Political attitudes have a significant explanatory effect in the Income_Random treatment (column 2). Subjects who self-identify as more right-leaning are less likely to choose the high tax rate. This is consistent with the expectation that left-leaning people prefer more redistribution than more liberal or right-leaning people.[52] I find no significant effect of political attitudes for the Baseline treatment.

[51] In the presence of moral hazard, I also included behavioral effects and self-statements in this regression.

[52] The distribution of stated political views in the Income_Random treatment is left skewed with a mass at the left center. It has the following parameters: skewness = 0.04 and kurtosis = 2.04.

Subjects who agree with the statement that “it is fair for talent to determine income inequality” (*talent_fair*) are less likely to choose the high tax in the BASELINE treatment.[53] On the other hand, subjects who agree with the statement that “no matter how much talent you have, you can always change it a bit” (*not_only_talent*) are more likely to choose the high tax rate in the BASELINE treatment, while they are less likely to in the INCOME_RANDOM treatment.

I also control for subjects’ age as well as their self-stated experience as a participant in experimental studies. Older subjects have a significant but only slightly higher probability of choosing the high tax rate in the BASELINE treatment. A higher experience as a participant in experimental studies, however, decreases the share of high tax rate choices significantly in this treatment. In the INCOME_RANDOM treatment, neither of these two variables has a significant effect on the high tax share. The effect of age on effort is significant, as older subjects solve slightly fewer puzzles. Although this effort is significant, it is not substantial in magnitude. More experience as a participant in experimental studies has no significant effect in the BASELINE treatment, while the effect is significant and substantial in the TAX_RANDOM treatment.

[53]Interestingly, these subjects also have a slightly higher effort provision. This is not significant for BASELINE, but it is significant for TAX_RANDOM.

Table C.5: Behavioral effects in rounds 2 – 6 (single) I.

	high tax	high tax	high tax	high tax	high tax	high tax	high tax	high tax
risk	0.000993							
	(0.000744)							
guess		0.00224*						
		(0.00133)						
keep			-0.0000347					
			(0.000155)					
altruism_wp13421				0.00636				
				(0.0198)				
altruism_wp13459					0.00000403			
					(0.0000949)			
overconfident						0.161***		
						(0.0386)		
underconfident							0.210***	
							(0.0471)	
confidence								0.0158
								(0.0201)
Constant	0.430***	0.396***	0.518***	0.447**	0.491***	0.482***	0.470***	0.467***
	(0.0794)	(0.0869)	(0.102)	(0.175)	(0.0516)	(0.0390)	(0.0372)	(0.0635)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Rounds	2-6	2-6	2-6	2-6	2-6	2-6	2-6	2-6
R^2	.0036825	.0110338	.0003514	.0008682	3.27e-06	.0056656	.016502	.0024514
Observations	480	480	480	480	480	480	480	480

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. More detailed information on the variables can be found in Table C.8.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.6: Behavioral effects in rounds 2 – 6 (single) II.

	high tax	high tax	high tax	high tax	high tax	high tax	high tax	high tax
politics	-0.00445							
	(0.0151)							
talent_fair		-0.0225*						
		(0.0117)						
not_only_talent			0.0236**					
			(0.0102)					
not_only_talent2				0.00767				
				(0.0107)				
age					0.00390			
					(0.00245)			
gender						-0.0578		
						(0.0741)		
business/economics							0.106**	
							(0.0453)	
num_experiments								-0.0748**
								(0.0336)
Constant	0.506***	0.612***	0.308***	0.449***	0.392***	0.571***	0.463***	0.681***
	(0.0559)	(0.0678)	(0.0990)	(0.0843)	(0.0817)	(0.116)	(0.0421)	(0.0528)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Rounds	2-6	2-6	2-6	2-6	2-6	2-6	2-6	2-6
R^2	.0003001	.0148332	.0077812	.0019097	.0030746	.0031305	.0089408	.0233098
Observations	480	480	480	480	465	480	480	480

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. More detailed information on the variables can be found in Table C.8.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.A.4 Individual strategies

Here, I use (behavioral) variables that might have an impact on subjects' choice of strategy. In different regressions, I test whether certain behavioral or demographic variables can explain a subject's played strategy. In other words, I analyze whether subjects who choose one of the most common strategies have certain behavioral or demographic characteristics in common.

I start with the variable $underconfident$. Table C.9 shows that subjects who were underconfident of their performance in the first round are over 30 percentage points more likely to play the "start high switch to low" strategy. For all other strategies, there is no significant effect of $underconfident$.

The experience as an experiment participant is captured in the variable *num_experiments*. Testing its impact on the probability of choosing different strategy profiles, I find that more experienced subjects are more likely to choose the "always low" strategy and they follow significantly more often the "if net payer in previous round, choose low tax; if net receiver in previous round, choose high tax" strategy principle. This is in line with the hypotheses and findings presented in Matthey and Regner (2013) who find that experienced subjects might be more selfish and Benndorf et al. (2017) who hypothesize that more experienced subjects are better at choosing payoff-maximizing actions. While the "always low" strategy can be interpreted as low redistributive preferences, the "if net payer in previous round, choose low tax; if net receiver in previous round, choose high tax" strategy principle can be seen as a payoff-maximizing (self-serving) action.

Table C.7: Behavioral effects in rounds 2 – 6 (combined).

	high tax	high tax	effort	effort
round	-0.0266**	-0.0696***	0.327***	0.167***
	(0.0112)	(0.0140)	(0.0302)	(0.0276)
risk	0.00113	0.000664	-0.0159*	-0.0101
	(0.000866)	(0.00103)	(0.00911)	(0.0106)
guess	0.00174	0.00184	0.00151	-0.00483
	(0.00160)	(0.00196)	(0.0111)	(0.0133)
keep	-0.0000122	-0.0000929	0.000130	-0.000403
	(0.000110)	(0.0000808)	(0.000858)	(0.00154)
altruism_wp13421	-0.0137	-0.0186	-0.0983	0.0194
	(0.0170)	(0.0262)	(0.107)	(0.191)
altruism_wp13459	-0.0000914	-0.0000219	0.000738	-0.00204
	(0.000133)	(0.000302)	(0.000644)	(0.00260)
overconfident	0.149	0.238	-1.297	-1.162**
	(0.140)	(0.196)	(0.996)	(0.576)
underconfident	0.216***	0.146*	0.790	1.560***
	(0.0691)	(0.0858)	(0.684)	(0.398)
confidence	0.00829	0.0304	-0.215	0.0145
	(0.0205)	(0.0203)	(0.159)	(0.195)
politics	0.00779	-0.0237	0.0519	0.268
	(0.0146)	(0.0196)	(0.111)	(0.229)
talent_fair	-0.0291***	0.0123	0.00991	0.116*
	(0.00986)	(0.0133)	(0.0834)	(0.0662)
not_only_talent	0.0209**	-0.0334**	-0.0661	-0.0337
	(0.0106)	(0.0136)	(0.116)	(0.135)
not_only_talent2	0.00395	0.00487	-0.0777	-0.0657
	(0.0160)	(0.0135)	(0.0886)	(0.0967)
age	0.00945**	-0.00179	-0.119***	-0.0924***
	(0.00432)	(0.00560)	(0.0349)	(0.0316)
gender	-0.108	0.0758	1.294	-0.246
	(0.0850)	(0.0551)	(0.848)	(0.749)
quantitative_study	0.127	-0.0755	-0.880	0.151
	(0.0830)	(0.114)	(0.577)	(0.985)
num_experiments	-0.0783*	-0.0346	0.550	1.097***
	(0.0400)	(0.0458)	(0.430)	(0.293)
Constant	0.546*	1.119***	16.12***	14.00***
	(0.302)	(0.339)	(1.945)	(3.103)
Treatment	BASELINE	INCOME_RANDOM	BASELINE	TAX_RANDOM
Rounds	2-6	2-6	2-6	2-6
R^2	.1170197	.1097908	.287289	.3226843
Observations	465	385	465	395

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *round* reflects the time trend. More detailed information on the variables can be found in Table C.8. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.8: Behavioral variables.

risk	takes values between 0 and 100 with a step size of 0.5. *risk* is equal to 0 for full risk aversion and is equal to 100 for risk neutrality.
guess	is the guessed number in a beauty contest and reflects a higher level-k, when *guess* decreases.
keep	is the amount kept in a dictator decision and yields less altruistic preferences, when *keep* increases:
*altruism_wp*13421	answers the question "How much would you be willing to give to a good cause without expecting anything in return?" on a scale from 0 ("not at all willing to do this" to 10 ("very willing to do this").
*altruism_wp*13459	answers the question "Imagine the following situation: Today you unexpectedly received 1000 euros. How much of the money would you donate to a good cause?".
overconfident	is equal to 1 if a subject stated that her performance was above average while it actually was below or equal to average performance.
underconfident	is equal to 1 if a subject stated that her performance was below average while it actually was above or equal to average performance.
confidence	reflects the difference in the believed and actual performance in eight quiz questions including versions of the cognitive reflection test (CRT).
politics	reflects the self-stated political view of each subject on a menu scale between 0 (left) and 9 (right).
talent_fair	presents the opinion for the statement "I think it's fair for talent to determine income inequality." on a scale from 0 ("Do not agree at all") to 10 ("agree completely").
not_only_talent	presents the opinion for the statement "No matter how much talent you have, you can always change it a bit." on a scale from 0 ("Do not agree at all") to 10 ("agree completely").
*not_only_talent*2	presents the opinion for the statement "You can always significantly change your own talents." on a scale from 0 ("Do not agree at all") to 10 ("agree completely").
age	presents the self-reported age of the subjects (Median age: 23).
gender	presents the self-reported gender of the subjects (female:57 percent, male: 42 percent, diverse: 1 percent)
business/economics	is equal to 1 if the studyfield is business administration or economics (28.13 percent).
num_experiments	reflects the self-reported number of experiments that a subject has participated in. This variable is categorical with 1 for no experience (0 previous experiments), 2 for 1 to 5 experiments, 3 for 6 to 10 experiments, and 4 for more than 10 experiments.

Table C.9: Individual behavioral analysis.

	Strategy					
	alwayslow	alwayshigh	high2low	low2high	net_0dev	net_1dev
underconfident	-0.0698**	0.0302	0.307*	-0.0930***	0.0302	-0.226**
	(0.0277)	(0.104)	(0.162)	(0.0310)	(0.103)	(0.114)
Constant	0.0698**	0.0698**	0.0930***	0.0930***	0.0698***	0.326***
	(0.0277)	(0.0287)	(0.0324)	(0.0310)	(0.0265)	(0.0517)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
R^2	.0077519	.0012617	.0803987	.0105708	.0012617	.0225233
Observations	96	96	96	96	96	96

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *underconfident* is equal to 1 if a subject stated that her performance was below average while it was actually above or equal to average performance. *underconfident* is elicited after the real-effort task in the first round of play. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C.10: Individual behavioral analysis.

	Strategy					
	alwayslow	alwayshigh	high2low	low2high	net_0dev	net_1dev
num_experiments	0.0582**	-0.0272	0.0463	0.0175	0.0529*	0.0160
	(0.0254)	(0.0200)	(0.0297)	(0.0278)	(0.0277)	(0.0494)
Constant	-0.0848*	0.142**	0.00782	0.0390	-0.0609	0.262**
	(0.0452)	(0.0687)	(0.0710)	(0.0705)	(0.0598)	(0.133)
Treatment	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
R^2	.0601189	.0114006	.0203911	.0041801	.0430194	.0012562
Observations	96	96	96	96	96	96

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *num_experiments* reflects the self-reported number of experiments that a subject has participated in. This variable is categorical with 1 for no experience (0 previous experiments), 2 for 1 to 5 experiments, 3 for 6 to 10 experiments, and 4 for more than 10 experiments. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.A.4.1 Individual strategies in Income_Random

Table C.11: Strategies in Income_Random.

Strategy	Income_Random
alternator	2
always low	4
always high	8
start low switch to high	9
start high switch to low	10
start and keep low except try high once	7
start and keep high except try low once	12
start and keep low except try high twice (separately)	1
start and keep high except try low twice (separately)	0
start and keep low except try high for 2 consecutive rounds	7
start and keep high except try low for consecutive 2 rounds	8

Notes: This table shows the number of subjects who applied different specified strategies in Income_Random.

3.A.5 Learning

Learning effects are not very strong in the real-effort task I chose. The effect is strongest between the first two rounds. Table C.12 shows the change from round one to round two for the two treatments Baseline and Tax_Random in columns 1 and 3. In both treatments the mean number of puzzles solved increases by more than one in the first round, but in the subsequent rounds this increase in puzzles solved flattens to 0.32 and 1.7 (see columns 2 and 4). The estimate for *tax_random* in column 5 shows that there is no significant difference in learning over time between the two treatments. Testing for marginal effects between certain rounds, I find positive learning effects until round 5, but no difference between rounds 5 and 6 in Baseline. For the Tax_Random treatment, I find no significant learning effect after round 3.[54]

[54]For Baseline the test results are: p-value=0.001 (round 3 to 4), p-value=0.054 (round 4 to 5), p-value=1.000 (round 5 to 6), and for Tax_Random the test results are: p-value=0.136 (round 3 to 4), p-value=0.417 (round 4 to 5), p-value=0.567 (round 5 to 6).

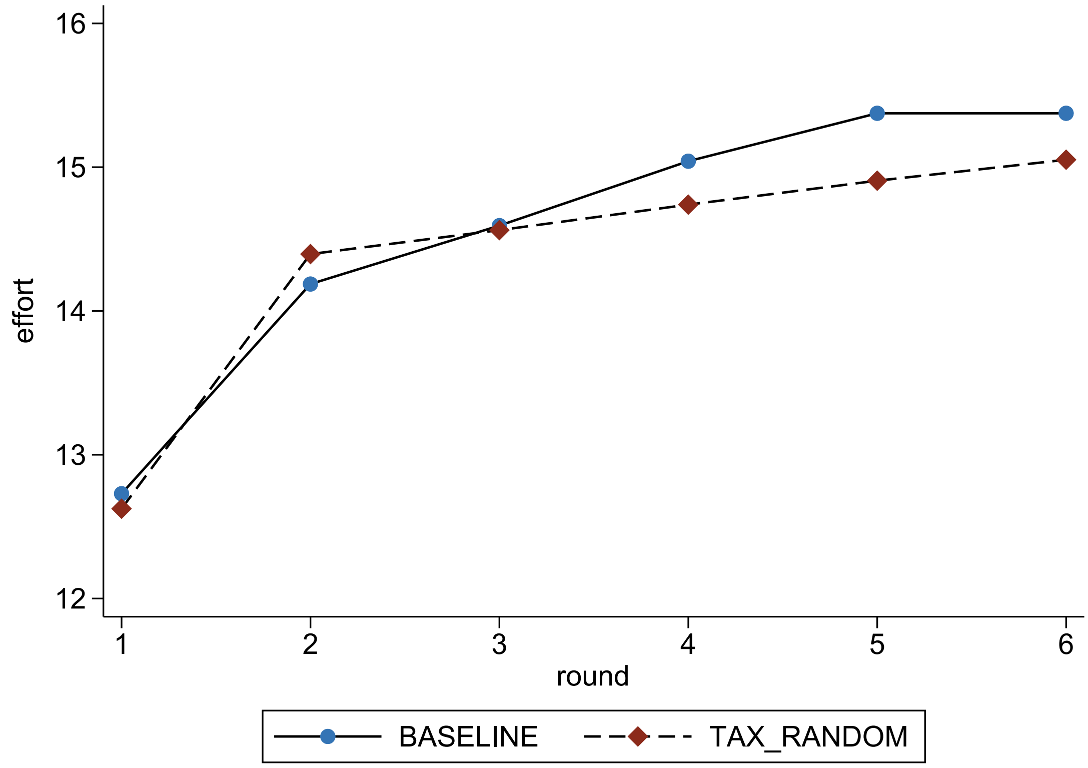
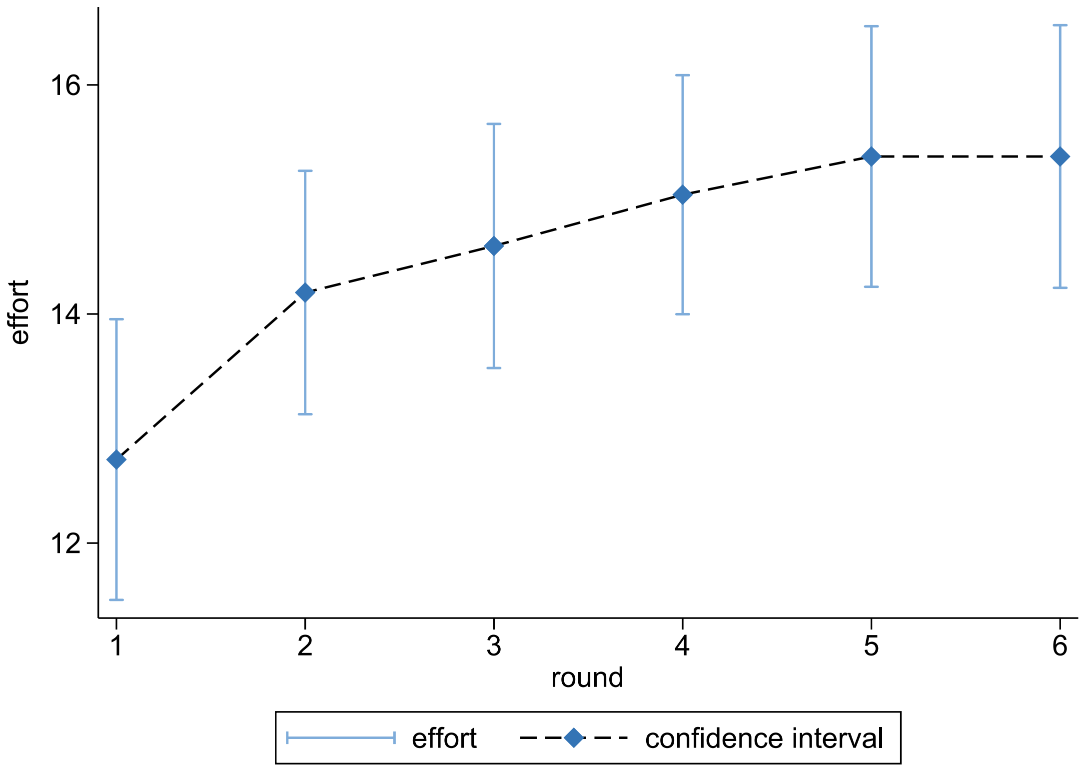


Figure C.5: Learning in the real-effort task.

Table C.12: Learning effect in real-effort task.

	effort	effort	effort	effort	effort
round	1.458***	0.316***	1.771***	0.166***	0.444***
	(0.126)	(0.0315)	(0.0856)	(0.0283)	(0.0232)
TAX_RANDOM					-0.170
					(0.504)
Constant	11.27***	13.65***	10.85***	14.07***	13.00***
	(0.514)	(0.373)	(0.369)	(0.249)	(0.430)
Treatment	BASELINE	BASELINE	TAX_RANDOM	TAX_RANDOM	BASELINE & TAX_RANDOM
Rounds	1-2	2-6	1-2	2-6	1-6
R^2	.0670865	.0268976	.1010221	.0083966	.0756097
Observations	192	480	192	480	1152

Notes: Bootstrapped standard errors (in parentheses), clustered at the session level. *round* reflects the time trend. TAX_RANDOM is equal to 1 for the TAX_RANDOM treatment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



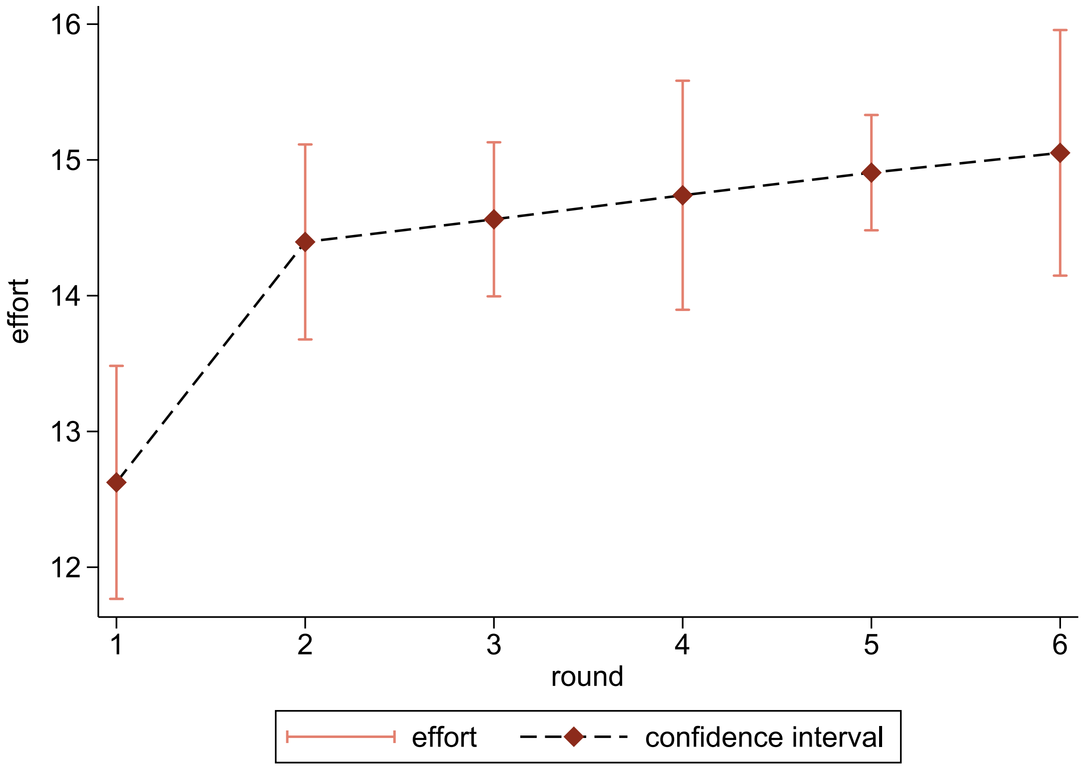


Figure C.6: Average effort provided over time in each treatment. BASELINE treatment on the left and TAX_RANDOM treatment on the right-hand side.

3.B Instructions

The following instructions were shown to participants on screen, but were also available as hard copies. The first part was shown prior to the first tax decision. After the first part of the instructions, I presented the four examples on screens to demonstrate the differences between the tax rates.[55] After the first tax decision, subjects read the instructions for the second part. I show the Baseline instructions here. The first part of the instructions are the same for Income_Random. Phrases used in the Tax_Random treatment that diverted from the Baseline instructions are in italic and initialized with TR.

In the second part, I again show the Baseline instructions which are basically the same for Tax_Random, apart from a difference in the summary of procedures' first bullet point saying that percentages were imposed and not chosen. Phrases used in the Income_Random treatment that diverted from the Baseline instructions are in italics and initialized with IR.

[55]Here I present the examples for the Baseline treatment. The shown numbers are the same for all treatments. For the Tax_Random treatment, I adjusted the wording of "tax rate choice" to "tax rate assignment" as in the first part of the instructions.

Part 1 Welcome! Thank you for participating in today's experiment.

Please read the following instructions for the experiment procedure carefully. If you have any questions please feel free to contact the experimenters. Please do not talk to your neighbors and remain quiet throughout the experiment. If you have a question, please raise your hand. We will then come to your seat and answer your question personally.
In this experiment, you can earn money through your decisions and personal effort. How much you earn depends on your decisions (or your effort) and on the decisions and effort of other participants randomly assigned to you. At the end of the experiment, you will definitely get your payoff.

All participants receive (and are reading) the same instructions.

You will remain completely anonymous for us and for the other participants. We do not store any data in connection with your name.

The Experiment

The experiment consists of three phases:

1. Choice of the percentage
   TR: *announcement of the percentage*
2. Effort and income
3. Application of the percentage

All phases are repeated a total of 6 times, so the experiment consists of 6 rounds.

1. Phase

You and also all other participants choose individually between two offered percentages. You can choose either 15 or 85. This choice is the same for everyone. You will then learn the number of other participants who have the same percentage.

TR: *You and also all other participants will be randomly assigned to one of two different percentages. These percentages are either 15 or 85. You will be told the percentage as well as the number of other participants with the same percentage.*

2. Phase

Income is generated. You will receive the concrete instructions for this phase after completion of the 1st phase.

3. Phase

The percentage is applied and thus determines your payoff. The exact effect is explained below.

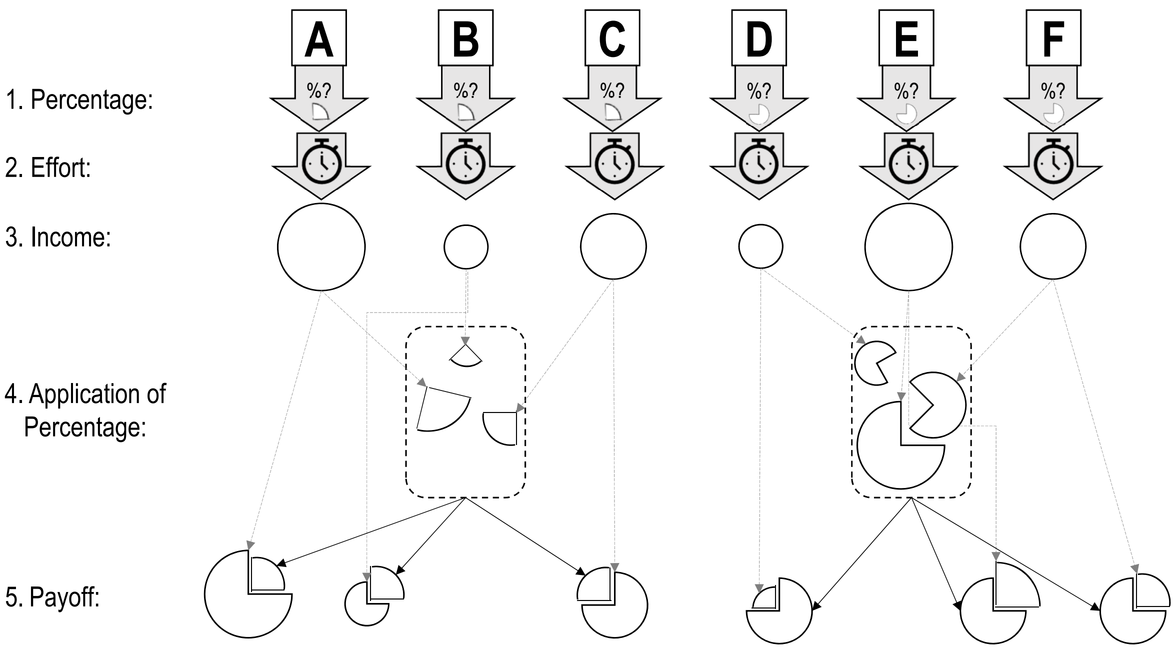
After that you will be given the information on how much you have earned in this period and what payoff you will get for the completed round. Your final payoff for the experiment is equal to the sum of your payoffs in each round. Your income as well as your payoff will be noted in points during the experiment. You will receive your final payoff in Euros at the end of the experiment. Here *100 points* correspond to one Euro or in other words:

1 Point = *0,01 Euro.*

Groups and application of the percentage

You will be randomly assigned to a group consisting of a total of *16* participants before the start of the actual experiment. These groups remain the same throughout the entire procedure. Your payoff depends on your decisions and effort, as well as on the decisions and effort of the other participants in your group.
Your payout in a round is co-determined by the percentage. This works as follows:
Income is generated.
The application of the percentage will first deduct a portion (corresponding to your previously randomly assigned percentage) from your income. This part is added

to the parts of those participants who have chosen the same percentage as you. Subsequently, you and the other participants who have chosen the same percentage as you will each receive equal payoffs of this initially deducted income.
TR: *This part is added to the parts of those participants who have been randomly assigned the same percentage as you. Subsequently, you and the other participants who were randomly assigned the same percentage as you will each receive this initially deducted income back in equal shares.*
The application of the percentage is illustrated by the following picture as an example:
(Participants A, B, and C as well as D, E, and F have each chosen the same percentage).
TR:*(Participants A, B, and C as well as D, E, and F have each been randomly assigned the same percentage assigned – the number of participants per percentage is known before the assignment.)*



The percentage is applied regardless of the number of participants who have chosen the same percentage. If no other participant has chosen the same percentage as you, you will get your initially deducted income paid off again.

TR: *The percentage is applied regardless of how many participants have been randomly assigned the same percentage. If no other participant has been randomly assigned the same percentage as you, you will be paid your initially deducted income again.*

Summary of the procedure

1. Percentage. You choose one of the two offered percentages

   TR: *You will be randomly assigned one of the two percentages.*

2. Effort. Income is generated (more details to follow separately)

3. Application. The percentage is applied, which partly determines your payoff for this round.

   (a) Applying the percentage will deduct a part of your earned income according to the selected percentage.

   (b) The deducted part of your income and the parts of all other group members who have chosen the same percentage will be summed up.

   TR: *The deducted part of your income and those of all other group members who were randomly assigned the same percentage are added up.*

   (c) By applying the percentage, this sum is paid off equally to the same participants.

4. Feedback. You will receive feedback on your earned income and payoff in this round.

Examples

Below you can see four examples of the application of the percentage.

After the examples you will find the detailed instructions again, in case you want to read again.

Example 1

Your generated income is 250 points:

A total of 8 participants choose 15.0% and 8 choose 85.0%.

Assume you have the percentage 15.0%.

You can see the following overview of the two percentages and their application:

(Each participant learns the number of other participants who have chosen the same percentage after his decision).

Overview	**15.0%**	**85.0%**
Number of participants	8	8
average income	219 points	219 points
Sum of deductions by the percentage	225 points	1275 points
individual payment through the application of the percentage	28 points	159 points

In this example, your final **payout** is **241 points**. Your generated **income was 250 points**. This payout results from the following calculation method:

Calculation of your payout

A total of 8 participants opted for the 15.0% percentage:

Your generated income	250 points
Your deduction by percentage 15.0%	37 points
Your direct remuneration (after deduction)	213 points
Sum of all deductions = 225 points ÷ 8 = 28 points*.	
Your indirect remuneration (payment deductions in equal parts)	28 points
Your final payout (total remuneration)	**241 points**

*Points are rounded up in your sense

Example 2

Your generated income is 250 points:

A total of 8 participants voted 15.0% and 8 voted 85.0%.

Assume you have the percentage 85.0%.

You can see the following overview of the two percentages and their application:

(Each participant learns the number of other participants who have chosen the same percentage after his decision).

Overview	**15.0%**	**85.0%**
Number of participants	8	8
average income	219 points	219 points
Sum of deductions by the percentage	225 points	1275 points
individual payment through the application of the percentage	28 points	159 points

In this example, your final **payout** is **197 points**. Your generated **income was 250 points**. This payout results from the following calculation method:

Calculation of your payout

A total of 8 participants opted for the 85.0% percentage:

Your generated income	250 points
Your deduction by percentage 85.0%	212 points
Your direct remuneration (after deduction)	38 points
Sum of all deductions = 1275 points ÷ 8 = 159 points*.	
Your indirect remuneration (payment deductions in equal parts)	159 points
Your final payout (total remuneration)	**197 points**

*Points are rounded up in your sense

Example 3

Your generated income is 250 points:

A total of 8 participants voted 15.0% and 8 voted 85.0%.

Assume you have the percentage 15.0%.

You can see the following overview of the two percentages and their application:

(Each participant learns the number of other participants who have chosen the same percentage after his decision).

Overview	15.0%	85.0%
Number of participants	8	8
average income	344 points	344 points
Sum of deductions by the percentage	375 points	2125 points
individual payment through the application of the percentage	47 points	266 points

In this example, your final **payout** is **260 points**. Your generated **income was 250 points**. This payout results from the following calculation method:

Calculation of your payout

A total of 8 participants opted for the 15.0% percentage:

Your generated income	250 points
Your deduction by percentage 15.0%	37 points
Your direct remuneration (after deduction)	213 points
Sum of all deductions = 375 points ÷ 8 = 47 points*.	
Your indirect remuneration (payment deductions in equal parts)	47 points
Your final payout (total remuneration)	**260 points**

*Points are rounded up in your sense

Example 4

Your generated income is 250 points:

A total of 8 participants voted 15.0% and 8 voted 85.0%.

Assume you have the percentage 85.0%.

You can see the following overview of the two percentages and their application:

(Each participant learns the number of other participants who have chosen the same percentage after his decision).

Overview	15.0%	85.0%
Number of participants	8	8
average income	344 points	344 points
Sum of deductions by the percentage	375 points	2125 points
individual payment through the application of the percentage	47 points	266 points

In this example, your final **payout** is **304 points**. Your generated **income was 250 points**. This payout results from the following calculation method:

Calculation of your payout

A total of 8 participants opted for the 85.0% percentage:

Your generated income	250 points
Your deduction by percentage 85.0%	212 points
Your direct remuneration (after deduction)	38 points
Sum of all deductions = 2125 points ÷ 8 = 266 points*.	
Your indirect remuneration (payment deductions in equal parts)	266 points
Your final payout (total remuneration)	**304 points**

*Points are rounded up in your sense

Here are a few short comprehension questions about the procedure of the experiment. Please click on "Next" for this.

Next

Part 2

Encryptio

These are the instructions for the second phase of the experiment.
In this phase you have to solve so-called encryptio. These are sequences of digits that are to be transferred into a sequence of letters with the help of a table in which a letter is assigned to each digit. The tables – and thus the assignment of the digits and letters – change with each encryptio.

Using the table, you transfer the sequence of digits shown into a sequence of letters and enter these letters into the field below the image. Afterwards you press "Senden." A new encryptio follows directly. The letters are not case-sensitive. So you can also enter lowercase letters.

For each correct entry you will receive *10* points. The reward for all correct entries is equal to the income you earn in this round. You have a total of *3* minutes in each round.
IR: *The number of solved encryptio has* ***nothing*** *to do with your income in this round.*[56] *Your income in this round is equal to the income that a random other participant earned in a past experiment by solving encryptio at the same time. Thereby, one correctly solved encryptio is equal to 10 points.*

Here you can see an example of an encryptio:

[56]This was also emphasized on the (wait) screens before the task.

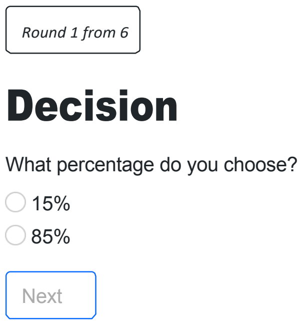
Above the picture: Please transfer the sequence of digits shown into the corresponding letter combination using the table.
Input field: Enter the sequence of letters

Summary of the procedure

1. Percentage. You choose one of the two offered percentages
   TR: *You will be randomly assigned one of the two percentages.*
2. Effort. Income is generated (more details to follow separately)
3. Application. The percentage is applied, which partly determines your payoff for this round.
   (a) Applying the percentage will deduct a part of your earned income according to the selected percentage.
   (b) The deducted part of your income and the parts of all other group members who have chosen the same percentage will be summed up.
   TR: *The deducted part of your income and those of all other group members who were randomly assigned the same percentage are added up.*
   (c) By applying the percentage, this sum will be payed out equally to the same participants.
4. Feedback. You will receive feedback on your earned income and payoff in this round.

3.C Experiment screenshots

Here, I present representations of the most important screens in my experiment, when the tax decision was made, what the task interface looked like, and an example of what the feedback page after the task looked like. These screenshots are taken from the Baseline treatment and have been translated as the experiment was originally conducted in German.



Encryptio

Time left for this page: **2:40**

Please transfer the sequence of digits shown into the corresponding letter combination using the table.

V	P	H	N	K	M	D	X	Q	C
5	1	3	8	2	4	9	7	0	6

85470

nvmxq | Send

Repeat the instructions for the encryptio:

Using the table, please transfer the sequence of digits shown into a sequence of letters and enter these letters into the button below the image and press "Send".

Afterwards a new Encryptio will be displayed directly. You will receive 10 points for each correct entry. The reward for all correct entries is equal to the income you earn in this round.

You have a total of 3 minutes for this in each round.

Summary of the procedure

1. **Percentage.** You choose one of the two offered percentages
2. **Stake.** Your income results from the number of correct entries (see above)
3. **Application.** The percentage is applied, which determines your payout for that round.
   1. By applying the percentage, a part of your earned income will be deducted from you according to the selected percentage.
   2. The deducted part of your income and the parts of all other group members who have chosen the same percentage are added up.
   3. By applying the percentage, this sum is paid equally to the same participants.
4. **Feedback.** You will receive feedback on your earned income and payout in this Round.

Round 1 from 6

Result

You have solved **9 Encryptio** correctly, earning **90 points of income in the** task. After applying the percentage, your **payout** is **50 points.**

Overview	**15.0%**	**85.0%**
Number of participants	8	8
average solved encryptio	7,75	4,25
Sum of deductions by the percentage	91 points	287 points
individual payment through the application of the percentage	11 points	36 points

Calculation of your payout

A total of 8 participants opted for the 85.0% percentage:

Your solved encryptio	9
Your earned income	90 points
Your deduction by percentage 85.0%	76 points
Your direct remuneration (after deduction)	14 points
Sum of all deductions = 287 points; 287 points ÷ 8 = 36 points *	
Your indirect remuneration (payment deductions in equal parts)	36 points
Your final payout (total remuneration)	**50 points**

*Points are rounded up in your sense

Next

Subjects' feedback on the experiment

Subjects' individual feedback on the experiment was very positive. Subjects had the chance to rate the experiment and to give a feedback after they had finished. On a scale from 0 to 5, the average rating was 4.53 (number of evaluations: 68, average payoff of evaluating subjects: 19.94). This rating was conducted in the DICELab only. Written feedback was also very positive and insightful. Subjects stated they had fun and felt challenged ("Was a lot of fun and awakened my ambition!" but also some shared thoughts about the experiment ("The distribution formula does not so much honor the personal contribution. The possibility to share I felt again as a nice gesture!" or"I'm not sure how much effort you put into the typing tasks as a participant if it doesn't determine your payout."). The statements were translated from German.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*, 488–500.

Anderson, S. P., A. De Palma, and B. Kreider (2001). Tax incidence in differentiated product oligopoly. *Journal of Public Economics 81*(2), 173–192.

Andreoni, J. and J. Miller (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Bartels, L. M. (2008). *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton University Press.

Basu, K. (1997). Why are so many goods priced to end in nine? And why this practice hurts the producers. *Economics Letters 54*(1), 41–44.

Benedek, D., R. A. De Mooij, M. Keen, and P. Wingender (2020). Varieties of VAT pass through. *International Tax and Public Finance 27*(4), 890–930.

Benndorf, V., C. Moellers, and H.-T. Normann (2017). Experienced vs. inexperienced participants in the lab: do they behave differently? *Journal of the Economic Science Association 3*(1), 12–25.

Benndorf, V., H. A. Rau, and C. Sölch (2019). Minimizing learning in repeated real-effort tasks. *Journal of Behavioral and Experimental Finance 22*, 239–248.

Benzarti, Y., D. Carloni, J. Harju, and T. Kosonen (2020). What goes up may not come down: Asymmetric incidence of value-added taxes. *Journal of Political Economy 128*(12), 44380–4474.

Blundell, R. (2009). Assessing the temporary VAT cut policy in the UK. *Fiscal Studies 30*(1), 31–38.

Bochet, O. and S. Siegenthaler (2021). Competition and price transparency in the market for lemons: Experimental evidence. *American Economic Journal: Microeconomics 13*, 113–140.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*, 117–120.

Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90*(1), 166–193.

Buchanan, J. M. (1976). *The Limits of Liberty: Between Anarchy and Leviathan.* University of Chicago Press.

Bundeskartellamt (2014, September). Sektoruntersuchung Lebensmitteleinzelhandel (SU LEH). B2-15/11.

Cabrales, A., R. Nagel, and J. V. Rodríguez Mora (2012). It is hobbes, not rousseau: an experiment on voting and redistribution. *Experimental Economics 15*, 278–308.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review 97*(3), 818–827.

Carbonnier, C. (2007). Who pays sales taxes? Evidence from French VAT reforms, 1987–1999. *Journal of Public Economics 91*(5-6), 1219–1229.

Carbonnier, C. (2008). Différence des ajustements de prix à des hausses ou baisses des taux de la TVA: un examen empirique à partir des réformes françaises de 1995 et 2000. *Economie et Statistique 413*(1), 3–20.

Cason, T. N. and L. Gangadharan (2002). Environmental labeling and incomplete consumer information in laboratory markets. *Journal of Environmental Economics and Management 43*(1), 113–134.

Chancel, L., T. Piketty, E. Saez, and G. Zucman (2022). *World Inequality Report 2022*. Harvard University Press.

Charness, G. and U. Gneezy (2010). Portfolio Choice And Risk Attitudes: An Experiment. *Economic Inquiry 48*(1), 133–146.

Charness, G. and M. Rabin (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics 117*(3), 817–869.

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

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

Cho, I.-K. and D. M. Kreps (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics 102*(2), 179–221.

Conlon, C. T. and N. L. Rao (2020). Discrete prices and the incidence and efficiency of excise taxes. *American Economic Journal: Economic Policy 12*(4), 111–43.

Cotropia, C. and K. Rozema (2018). Who benefits from repealing tampon taxes? Empirical evidence from New Jersey. *Journal of Empirical Legal Studies 15*(3), 620–647.

Crossley, T. F., H. Low, and M. Wakefield (2009). The economics of a temporary VAT cut. *Fiscal Studies 30*(1), 3–16.

D'Acunto, F., D. Hoang, and M. Weber (2022, April). Managing households' expectations with unconventional policies. *The Review of Financial Studies 35*(4), 1597–1642.

Dalamagas, B. and S. Kotsios (2008). Personal income tax: Incentive or disincentive to work effort? *Revue économique 59*(4), 777–811.

Danz, D., N. Gupta, M. Lepper, L. Vesterlund, and K. P. Winichakul (2021). Going Virtual: A Step-by-Step Guide to Taking the In-Person Experimental Lab Online. *Economics: Faculty Publications, Smith College, Northampton, MA*.

Deffains, B., R. Espinosa, and C. Thöni (2016). Political self-serving bias and redistribution. *Journal of Public Economics 134*, 67–74.

Dickson, P. R. and A. G. Sawyer (1990). The price knowledge and search of supermarket shoppers. *Journal of Marketing 54*(3), 42–53.

Dovern, J., J. Frank, A. Glas, L. S. Müller, and D. Perico Ortiz (2022). Estimating pass-through rates for the 2022 tax reduction on fuel prices in Germany. *CESifo Working Paper No. 9963*.

Doyle Jr, J. J. and K. Samphantharak (2008). $2.00 Gas! Studying the effects of a gas tax moratorium. *Journal of Public Economics 92*(3-4), 869–884.

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

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

Engelmann, D., E. Janeba, L. Mechtenberg, and N. Wehrhöfer (2023). Preferences over taxation of high-income individuals: Evidence from a survey experiment. *European Economic Review 157*, 104505.

Engelmann, D. and M. Strobel (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review 94*(4), 857–869.

Espinosa, R., B. Deffains, and C. Thöni (2020). Debiasing preferences over redistribution: An experiment. *Social Choice and Welfare 55*(4), 823–843.

Evanschitzky, H., P. Kenning, and V. Vogel (2004). Consumer price knowledge in the German retail market. *Journal of Product & Brand Management 13*(6), 390–405.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Frohlich, N. and J. A. Oppenheimer (1990). Choosing justice in experimental democracies with production. *American Political Science Review 84*(2), 461–477.

Frohlich, N., J. A. Oppenheimer, and C. L. Eavey (1987). Laboratory results on rawls's distributive justice. *British Journal of Political Science 17*(1), 1–21.

Fuest, C., F. Neumeier, and D. Stöhlker (2021). The pass-through of temporary VAT rate cuts: Evidence from German supermarket retail. *CESifo Working Paper No. 9149*.

Genakos, C. and M. Pagliero (2022). Competition and pass-through: Evidence from isolated markets. *American Economic Journal: Applied Economics 14*(4), 35–57.

Gerber, A., A. Nicklisch, and S. Voigt (2019). The role of ignorance in the emergence of redistribution. *Journal of Economic Behavior & Organization 163*, 239–261.

Ginzberg, E. (1936). Customary prices. *American Economic Review 26*(2), 296–296.

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

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2006). The competitive advantage of sanctioning institutions. *Science 312*(5770), 108–111.

Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2014). On cooperation in open communities. *Journal of Public Economics 120*, 220–230.

Harsanyi, J. C. (1975). Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory. *The American Political Science Review 69*(2), 594–606.

Haucap, J., U. Heimeshoff, G. J. Klein, D. Rickert, and C. Wey (2021). Vertical relations, pass-through, and market definition: Evidence from grocery retailing. *International Journal of Industrial Organization 74*, 102693.

Haucap, J. and A. Rasch (2019). *Ökonomische Aspekte der Novellierung der HwO 2004*. Duesseldorf Institute for Competition Economics (DICE), University of Duesseldorf.

Henze, B., F. Schuett, and J. P. Sluijs (2015). Transparency in markets for experience goods: Experimental evidence. *Economic Inquiry 53*, 640–659.

Hindriks, J. and V. Serse (2022). The incidence of VAT reforms in electricity markets: Evidence from Belgium. *International Journal of Industrial Organization 80*, 102809.

Höchtl, W., R. Sausgruber, and J.-R. Tyran (2012). Inequality aversion and voting on redistribution. *European Economic Review 56*(7), 1406–1421.

Holt, C. and R. Sherman (1990). Advertising and product quality in posted-offer experiments. *Economic Inquiry 28*, 39–56.

Huck, S., G. K. Lünser, and J.-R. Tyran (2012). Competition fosters trust. *Games and Economic Behavior 76*, 195–209.

Huck, S., G. K. Lünser, and J.-R. Tyran (2016). Price competition and reputation in markets for experience goods: an experimental study. *RAND Journal of Economics 47*, 99–117.

Jaworski, T. and B. J. Wilson (2013). Go west young man: Self-selection and endogenous property rights. *Southern Economic Journal 79*(4), 886–904.

Jiménez-Jiménez, N., E. Molis, and Á. Solano-García (2020). The effect of initial inequality on meritocracy: A voting experiment on tax redistribution. *Journal of Economic Behavior & Organization 175*, 380–394.

Kaas, K. P. (1993). Symbiotic relationships between producers and retailers in the German food market? *Journal of Institutional and Theoretical Economics (JITE) 149*(4), 741–747.

Kleiner, M. M. (2000). Occupational licensing. *Journal of Economic Perspectives 14*, 189–202.

Kleiner, M. M. (2017). The influence of occupational licensing and regulation. *IZA World of Labor*, Article 392.

Kleiner, M. M. and A. B. Krueger (2010). The prevalence and effects of occupational licensing. *British Journal of Industrial Relations 48*, 676–687.

Kleven, H., C. Landais, M. Munoz, and S. Stantcheva (2020). Taxation and migration: Evidence and policy implications. *Journal of Economic Perspectives 34*(2), 119–142.

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

Kosonen, T. (2015). More and cheaper haircuts after VAT cut? On the efficiency and incidence of service sector consumption taxes. *Journal of Public Economics 131*, 87–100.

Leland, H. E. (1979). Quacks, lemons, and licensing: A theory of minimum quality standards. *Journal of Political Economy 87*, 1328–1346.

Li, J., S. Leider, D. Beil, and I. Duenyas (2021). Running online experiments using web-conferencing software. *Journal of the Economic Science Association 7*, 167–183.

Loy, J.-P., C. Ceynowa, and L. Kuhn (2020). Price recall: Brand and store type differences. *Journal of Retailing and Consumer Services 53*, 101990.

Lynch, M., R. M. Miller, C. R. Plott, and R. Porter (1986). Product quality, consumer information and “lemons” in experimental markets. In P. Ippolito and

D. Scheffman (Eds.), *Empirical Approaches to Consumer Protection Economics*, pp. 251–306. Washington, DC: Federal Trade Commission.

Mago, S. D. (2010). Costly buyer search in a differentiated products model: An experimental study. *Southern Economic Journal 77*, 282–312.

Mann, H. B. and D. R. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

Martínez, I. Z. (2022). Mobility responses to the establishment of a residential tax haven: Evidence from Switzerland. *Journal of Urban Economics 129*, 103441.

Matthey, A. and T. Regner (2013). On the independence of history: experience spill-overs between experiments. *Theory and decision 75*, 403–419.

Meltzer, A. H. and S. F. Richard (1981). A rational theory of the size of government. *Journal of Political Economy 89*(5), 914–927.

Miklos-Thal, J. and G. Shaffer (2021). Pass-through as an economic tool: On exogenous competition, social incidence, and price discrimination. *Journal of Political Economy 129*(1), 323–335.

Mill, W. and C. Schneider (2023). The bright side of tax evasion. *CESifo Working Paper No. 9149*.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2022). Managing self-confidence: Theory and experimental evidence. *Management Science 68*(11), 7793–7817.

Montag, F., A. Sagimuldina, and M. Schnitzer (2021). Does tax policy work when consumers have imperfect price information? Theory and evidence. *CESifo Working Paper No. 9138*.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review 85*(5), 1313–1326.

Nagle, T. T. and R. K. Holden (1987). *The Strategy and Tactics of Pricing: A Guide to Profitable Decision Making.* Prentice Hall Englewood Cliffs, NJ.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy 78*, 311–329.

Nozick, R. (1974). *Anarchy, State, and Utopia.* Basic Books.

Okun, A. M. (1975). *Equality and Efficiency: The Big Tradeoff.* Brookings Institution Press.

Rawls, J. (1971). *A Theory of Justice.* Harvard Books.

Recker, C. (2018). *Nachwuchssorgen im Handwerk.* Der Ordnungspolitische Kommentar Nr. 09/2018. Institut für Wirtschaftspolitik, University of Cologne.

Riordan, M. H. (1986). Monopolistic competition with experience goods. *The Quarterly Journal of Economics 101*, 265–279.

Röber, W. B. (2009). *Eine informationsökonomische Analyse des Handwerks.* PhD thesis, University of Lueneburg.

Rossouw, L., H. Ross, et al. (2020). An economic assessment of menstrual hygiene product tax cuts. *Gates Open Res 4*(137), 137. (not peer-reviewed).

Rüll, D. (2020). How to abolish indirect taxes on menstrual hygiene products. *CBS LAW Research Paper No. 20 - 28*.

Sarfati, H. (2015). OECD. In it together: Why less inequality benefits all. Paris, 2015. 332 pp. ISBN 978-9264-23266-2. *International Social Security Review 68*(4), 115–117.

Schildberg-Hörisch, H. (2010). Is the veil of ignorance only a concept about risk? An experiment. *Journal of Public Economics 94*(11-12), 1062–1066.

Schindler, R. M. and T. M. Kibarian (1996). Image communicated by the use of 99 endings in advertised prices. *Journal of Retailing 72*(2), 187–199.

Schindler, R. M. and P. N. Kirby (1997). Patterns of rightmost digits used in advertised prices: Implications for nine-ending effects. *Journal of Consumer Research 24*(2), 192–201.

Schmidheiny, K. and M. Slotwinski (2018). Tax-induced mobility: Evidence from a foreigners' tax scheme in Switzerland. *Journal of Public Economics 167*, 293–324.

Shaked, A. and J. Sutton (1982). Relaxing price competition through product differentiation. *Review of Economic Studies 49*, 3–13.

Stantcheva, S. (2021). Why Do We Not Support More Redistribution? New Explanations from Economics Research. In *Combating Inequality: Rethinking Government's Role*. The MIT Press.

Statistisches Bundesamt (Destatis) (2022). Genesis-Online. Data licence by-2-0 - for details see https://www.govdata.de/dl-de/by-2-0; own calculation and representation using Table 61111-0006; Codes: CC13-1213222100, CC13-1213221100, CC13-1213229100.

Stiglitz, J. E. (2012). *The Price of Inequality: How Today's Divided Society Endangers our Future*. WW Norton & Company.

Stiving, M. and R. S. Winer (1997). An empirical analysis of price endings with scanner data. *Journal of Consumer Research 24*(1), 57–67.

Strulov-Shlain, A. (2022). More Than a Penny's Worth: Left-Digit Bias and Firm Pricing. *The Review of Economic Studies 90*(5), 2612–2645.

Sutter, M. and H. Weck-Hannemann (2003). Taxation and the veil of ignorance–a real effort experiment on the laffer curve. *Public Choice 115*(1-2), 217–240.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy 64*(5), 416–424.

Timmons, E. J. and R. J. Thornton (2010). The licensing of barbers in the USA. *British Journal of Industrial Relations 48*, 740–757.

Timmons, E. J. and R. J. Thornton (2019). There and back again: The de-licensing and re-licensing of barbers in Alabama. *British Journal of Industrial Relations 57*, 764–790.

Tyran, J.-R. and R. Sausgruber (2006). A little fairness may induce a lot of redistribution in democracy. *European Economic Review 50*(2), 469–485.

Vanhuele, M. and X. Drèze (2002). Measuring the price knowledge shoppers bring to the store. *Journal of Marketing 66*(4), 72–85.

Vanhuele, M., G. Laurent, and X. Dreze (2006). Consumers' immediate memory for prices. *Journal of Consumer Research 33*(2), 163–172.

Weyl, E. G. and M. Fabinger (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, 80–83.

Statement of contributions

Statement of contribution

My co-author Justus Haucap contributed to the chapter

"VAT Pass-Through: The Case of a Large and Permanent Reduction in the Market for Menstrual Hygiene Products"

of my dissertation

"Essays in Applied Microeconomics."

All authors contributed *equally* to this chapter.

Signature of Justus Haucap ____________________

Statement of contribution

My co-authors Alexander Rasch, Nicolas Fugger, and Carina Fugger contributed to the chapter

"Occupational Licensing and Quality in a Market for Experience Goods: Experimental Evidence"

of my dissertation

"Essays in Applied Microeconomics."

Theory:	Alexander Rasch, Nicolas Fugger, Alisa Frey
Experiment:	Alisa Frey, Carina Fugger
Empirical Analysis:	Alisa Frey

Signature of Alexander Rasch ________________

Signature of Nicolas Fugger Nicolas Fugger

Signature of Carina Fugger C. Fugger

Eidesstattliche Versicherung

Ich, Frau Alisa Frey, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“erstellt worden ist.

Düsseldorf, den 24. Mai 2024

Unterschrift: ______________________