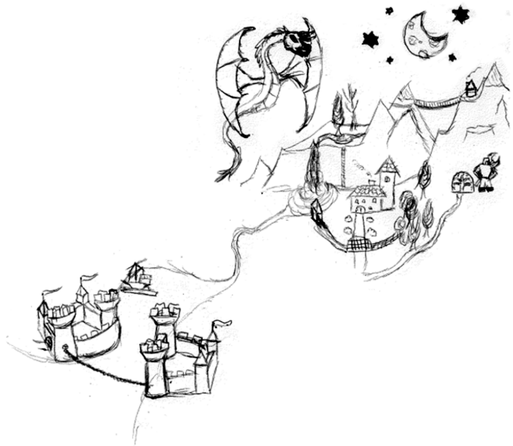
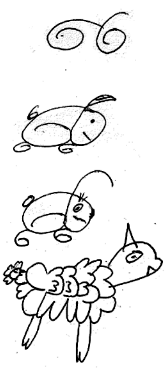
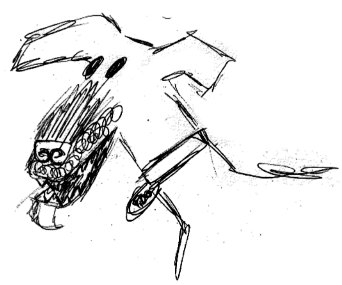
TEAM BEHAVIOR IN AUCTIONS, RECIPROCITY ON MARKETS, AND GENDER BIAS IN STUDENT EVALUATIONS OF TEACHING: ESSAYS IN EXPERIMENTAL ECONOMICS

MARIA KARMELIUK







DISSERTATION

TEAM BEHAVIOR IN AUCTIONS,
RECIPROCITY ON MARKETS, AND
GENDER BIAS IN STUDENT EVALUATIONS OF TEACHING:
ESSAYS IN EXPERIMENTAL ECONOMICS

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität München

2020

vorgelegt von
Maria Karmeliuk

Referent:	Prof. Dr. Martin G. Kocher
Koreferent:	Prof. Dr. Florian Englmaier
Promotionsabschlussberatung:	03.02.2021
Berichterstatter:	Prof. Dr. Martin G. Kocher
	Prof. Dr. Florian Englmaier
	Prof. Dr. Simeon Schudy
Mündliche Prüfung:	25.01.2021

To my parents Svetlana and Viktor,
without whom I would not have started my PhD,
to my husband Andrey,
without whom I would not have finished it,
to my son Erik,
who gave me a nice long break from it,
and all those
who contributed to this dissertation and will never read it

Acknowledgements

This dissertation hinged on the support of many people and organizations. I am grateful to everyone who helped me with my research as well as non-research related activities during my PhD.

First, I would like to thank my supervisor, coauthor and co-teacher Martin Kocher. Martin guided me through the maze of experimental research, always being supportive and available for questions and feedback on research projects. He made a great team lead at our chair and shaped our warm work environment. I am also thankful to my second supervisor and coauthor Florian Englmaier. Florian has helped me greatly with his feedback and conversations about research and academia. He uplifted me with his positive vibes even when I got a null result. Martin and Florian encouraged me to stick to the timeline when I did not believe myself capable of it (the result of these efforts is presented in the following chapters).

Further, I want to thank Simeon Schudy for agreeing to join my dissertation committee. Simeon is a wealth of knowledge in many areas of research, always open for discussion and brainstorming. Thanks for the tons of helpful feedback, for being the life of the party, for making our chair hold together and for the constant support.

I thank my coauthors Katharina Schüßler, Markus Reisinger, Artem Duplinskiy and Francisco Blasques for working with me on our projects. Special thanks go to Katharina who took over the experimental sessions in the laboratory while I was parenting my one-year old son. I am also grateful to the participants of the EEA Congress 2016, EBE Summer Meeting 2016, MGSE Colloquium, Behavioral Brown Bag seminar, Research Cake sessions and Micro Theory reading group for their invaluable thoughts, discussions and comments.

I am especially grateful to my colleagues at the chair of behavioral and experimental economics for being the best team I have ever worked with. My crew has been there to help, encourage, discuss, provide constructive criticism, solve problems, work hard and have fun together. Martin, Simeon, Lisa, Marvin, Felix, Stefan, Konstantin, Nadja, David, Gökhan, Kristina, Andrej, Clarissa, thank you for being a part of it. I am also thankful to everyone at Munich Graduate School of Economics who made my time here enjoyable, and to Katharina and Rainer for their support.

As a foreign employee who changed her last name and went on parental leave, I have had administrative difficulties here and there, which have been promptly solved by our administrative angels Silke Englmaier, Julia Zimmermann, Manuela Firmin and Angelica Schmidt. I am likewise grateful to the research assistants at the chair of behavioral and experimental economics and the MELESSA laboratory for providing excellent assistance to my experiments. I also thank the chair of behavioral and experimental economics, the chair of organizational economics, the Elite Network of Bavaria and the German Science Foundation (grant CRC TRR 190) for the generous funding of my research.

I want to thank my students, whom I taught and supervised during my stay at the MGSE, for asking tough questions, providing feedback and making me a better teacher.

The pictures on the front cover of this dissertation were drawn by the anonymous participants of my experiments at the MELESSA lab.

Finally, I am deeply indebted to my family and friends for their endless support. Only because of you I was able to start, carry out and finish this big little project. Thank you so much for having my back.

Maria Karmeliuk,

Munich, September 2020

Contents

List of Figures

List of Tables

Preface

The discipline of behavioral economics has challenged the standard assumptions behind the neoclassical *Homo Oeconomicus* – an unboundedly rational and selfish decision maker. The ideas that human behavior in economic situations can be driven by other forces than rationality and optimizing are not new to the economic literature. In fact, they appeared in the economic thought as early as the work of Adam Smith, Irving Fisher and John Maynard Keynes (see Ashraf et al., 2005, Pech and Milan, 2009, Thaler, 1997). Starting in the early 1950s, an American economist and cognitive psychologist Herbert Simon has planted the seeds for the field of bevavioral economics with his work on bounded rationality and heuristic decision making. The deviations from rational choice theory were further investigated and incorporated into choice models in the 1970s by two psychologists, Daniel Kahneman and Amos Tversky. They popularized the ideas of an "irrational" decision maker and are considered to be the fathers of modern behavioral economics. Since then, the field has grown and been widely recognized in research, teaching and policy advice.

Nowadays, behavioral economics does not conflict with neoclassical economics as much, but rather provides the latter with additional insights and serves as a part of all economists' toolkit (see Chetty, 2015). In fact, Levine (2012) asserts that due to this symbiosis behavioral economics "is doomed" as a standalone research field. On the other hand, some researchers criticize behavioral economics for the lack of empirical realism in modeling the decision making processes and being too close to the neoclassical "as-if" paradigm (Berg and Gigerenzer, 2010). With hundreds of discovered behavioral biases, behavioral economists are accused of oftentimes seeing a bias when there is none (Gigerenzer, 2018). Thus, an important point on the current research agenda of a behavioral economist is not detecting yet another deviation from selfish-rational behavior, but rather investigating in which contexts it plays an important role, how it affects the well-being of the involved parties, and what decision making processes stand behind this behavior.

This dissertation describes three behavioral aspects in three different contexts. Chapter 1 highlights the differences in the bidding behavior of individuals and teams in different types of auctions. Although neoclassical economics typically models decision makers as individuals, it has long been acknowledged in social sciences that people can behave differently when they make decisions in teams (refer to Kugler et al., 2012, for a review of the literature). One of the contexts where team behavior is widespread is bidding on both traditional brick-and-mortar and online auctions. Chapter 1 analyzes team and individual bidding behavior in a series of experimental auctions of most common formats: first- and second-price sealed-bid auctions and ascending-price (English) auctions.

The next behavioral aspect discussed in this dissertation is social preferences – one of the most prominent concepts in behavioral economics. It has been shown to matter in an ample spectrum of contexts, both in the lab and in the field (Fehr and Fischbacher, 2002, List, 2009). The study presented in Chapters 2 and 3 applies social preferences to markets and studies third-degree price discrimination when consumers have reciprocal preferences. According to traditional microeconomic theory, third-degree price discrimination – the practice of selling the same product at different prices to different customers – is always better than uniform pricing since it allows the firm to extract more consumer surplus. However, in practice this pricing policy can backfire: consumers tend to become angry when they find out that they pay higher prices than somebody else. Chapter 2 introduces a model of a monopolistic market where consumers have preferences for reciprocity and describes optimal third-degree price discrimination schemes under these conditions. The model is complemented by a laboratory experiment in Chapter 3, which tested reciprocal reactions to price discrimination and allowed to assess the optimal pricing strategy empirically.

Finally, Chapter 4 proposes an experimental design for the identification of gender bias in student evaluations of teaching. Gender differences were empirically documented in numerous economic contexts. A broad strand of literature claims that these differences can be partly attributed to gender discrimination, or gender bias – the propensity of a decision maker to perceive men as superior to women, or vice versa, even if their observable characteristics are comparable. In particular, gender bias is claimed to affect student evaluations of teachers (Boring, 2017, Mengel et al., 2019). These evaluations often influence hiring and promotion decisions at universities and thus indirectly affect the academic career of the faculty. Chapter 4 describes an experiment that varies the gender identity of the teacher and the quality of teaching exogenously, thereby allowing to identify how the gender bias shapes the differences in student evaluations, if at all. On top of that, the design includes a non-suggestive intervention to mitigate potential gender bias on the side of the students.

All three studies rely on the same methodology – laboratory experiments. Though frequently criticized for the lack of external validity, laboratory experiments offer a unique opportunity of tight control over the treatment variations, as well as beliefs and incentives of the participants. In Chapter 1, the laboratory environment allowes for the valuations of the bidders to be fixed and the differences between individual and team decision making processes to be minimized. It provides the only possible way to impose payoff functions and identify the optimal pricing strategy in Chapter 3. It makes it possible in Chapter 4 to vary the gender of the teacher without deception, minimizing the differences in teaching style and ensuring that the subjects trust that their feedback will reach the teachers. Being an indispensable tool to identify causality and disentangle different mechanisms of decision making, laboratory experiments constitute the methodological core of this dissertation. The detailed summary of the four chapters is provided below.

Two Heads are Better than One: Teams and Individuals in Standard Auction Formats. Chapter 1 is joint work with Martin Kocher. In this experiment, we study the differences in individual and team bidding behavior. It has been shown that teams differ from individuals on a variety of dimensions, including risk-taking, the ability to solve complex problems and moral behavior (Shupp and Williams, 2008, Maciejovsky et al., 2013, Kocher et al., 2018). In the context of auctions, it is often small teams of experts rather than individuals who decide on the bidding strategy (including oil and gas lease auctions, spectrum autions, etc.). However, empirical evidence on team bidding behavior is scarce and delivers mixed results. Our experiment was the first to compare teams and individuals in most common auction formats: first-price and second-price sealed-bid auctions, and English auctions.

In a laboratory experiment, our subjects participated in a series of repeated auctions with independent private values: either first-price or English auctions. Depending on the treatment, subjects either bid individually or in teams of three, with the possibility to communicate anonymously in a chat. The members of a team had the same valuations and thus the same amount of information and the same incentives as the individual bidders. After the repeated auctions ended, everyone participated in one second-price auction (we considered the behavior in this round to be a proxy for bidder's rationality). In addition, we elicited risk preferences in an investment task and retrieved a measure of cognitive abilities with the help of Raven's progressive matrices test. All parts of the experiment were incentivized. During the experiment we videotaped the subjects in order to analyze their facial expressions and estimate the effect of emotions on bidding behavior.

We found that in first-price auctions teams achieved higher profits than individuals, but there was no difference in English auctions. In our setting with independent private values, the optimal bidding strategy in English auction was very simple – staying in the auction until the current bid hits the valuation threshold. Both individuals and teams coped with this strategy equally well. In contrast, the optimal bids in first-price auctions are not that straightforward since they involve risk preferences and beliefs about the bidding strategies of the other bidders on the market. We observed that individuals overbid in comparison to teams and, as a result, achieved lower profits. It suggests that teams were more successful in understanding the complex optimal bidding strategy. Our speculation is corroborated by the correlation between the bids and proxies of rational behavior, such as bids in second-price auctions and scores on the Raven's test. This finding is in line with the literature asserting that teams are more rational decision makers than individuals (Maciejovsky et al., 2013).

In addition, we were able to explore the role of emotions in auctions. We observed that in first-price auctions subjects in teams experienced on average more positive emotions than subjects who bid individually. Naturally, winning an auction was associated with more positive and less negative emotions. Interestingly, emotions also had an effect on the future bids: sadness made subjects bid less in the next auction, while the feeling of fear increased their bids.

Optimal third-degree price discrimination and reciprocity: A theoretical model. From every textbook in microeconomic theory, an economics undergraduate learns that a monopolist always prefers to practice price discrimination whenever it is possible since this way he can effectively redistribute consumer surplus. However, there is ample evidence that consumers are disturbed by third-degree price discrimination when they become aware of it (Englmaier et al., 2012, Leibbrandt, 2020). Nevertheless, to this day theoretical literature incorporating behavioral reactions of consumers to prices is scarce.

In Chapter 2 I developed a model of a monopolistic market with two consumers who have preferences for reciprocity by applying the behavioral model of Falk and Fischbacher (2006). The consumers have linear demands[1] satisfying Spence-Mirrlees conditions, thereby one consumer always has a higher willingness-to-pay and is better off, or "richer" than the other consumer. Both consumers evaluate their well-being relative to a reference point, which includes both the price charged to the other consumer and his material payoff. If the consumer's actual payoff falls below his reference payoff, he considers the firm as unkind and is willing to punish it by lowering his demand. Analogously, if the consumer's payoff exceeds the reference payoff, he is willing to reward the firm by buying more. Each consumer puts certain weights on the price and the payoff of the other consumer in his reference point; this allowed me to model a situation when a lower price for the "poorer" consumer was deemed acceptable by the "richer" consumer.

My model has a unique equilibrium in a sensible range of parameters. It predicts that preferences for reciprocity put a constraint on price discrimination. The attractiveness of price discrimination decreases in the magnitude of reciprocal preferences of the "richer" consumer: the stronger he reacts to differential pricing, the less profitable it is to charge a higher price for him, with uniform pricing becoming optimal at some point. This reaction is mitigated if the consumer puts more weight on the "poorer" consumer's payoff when evaluating his reference point. On the other hand, positive reciprocity of the "poorer" consumer positively affects the profitability of price discrimination. Finally, price discrimination is more profitable when consumers differ strongly in their willingness-to-pay. When consumers are similar to each other, the benefits of price discrimination do not outweigh the costs imposed by the negative reciprocal reactions.

Optimal third-degree price discrimination and reciprocity: Experimental evidence. Chapter 3 is joint work with Florian Englmaier, Markus Reisinger and Katharina Schüßler. It introduces a laboratory experiment designed to test the predictions of the model in Chapter 2 and estimate its parameters. We simulated a monopolistic market by matching the subjects into groups of three consisting of a firm and two consumers with different payoff structures. The firm had a choice between a price

[1] According to neoclassical approach without reciprocal preferences.

discrimination scheme and a uniform price, and the consumers decided how much of a virtual good they buy from the firm. The market interaction was repeated for 20 periods, allowing us to structurally estimate the model for every individual. We had three treatments in our experiment. *No Info* treatment served as a control group: the subjects in the role of consumers learned neither the price nor the payoff of the other consumer on the market. In *Part Info* treatment we provided the consumers with the information about the price charged to the other consumer on the market, but not his payoff; and in *Full Info* treatment consumers learned both the price and the payoff of the other consumer. Under plausible assumptions the model predicted largest negative reciprocity and therefore lowest price discrimination rates in *Part Info* treatment, followed by *Full Info* treatment, followed by the highest frequency of price discrimination in *No Info* treatment.

Somewhat surprisingly, our results could not replicate consumer antagonism to price discrimination reported in the previous literature. We found no treatment differences in consumer demands, all of them being rather close to the selfish-rational benchmark with only minor deviations. The firms anticipated the consumer behavior accordingly: their pricing choices were not different from the best response to actual consumer demands in all treatments. The estimated parameters of the model were also very close to the neoclassical theory: the reciprocity parameters of the consumers were estimated to be very close to zero. The small deviations from selfish-rational behavior could not be captured by the model and were absorbed by the error term. We speculate that this result is a side effect of our experimental design. Repeated market interactions and the strategy method might have induced our subjects to make more rational decisions instead of producing emotional reactions to price discrimination. We also discuss the possibility that the reference price in our experiment was shaped by self/self comparisons rather than by self/other comparisons.

The determinants of gender bias in student evaluations of teaching. Chapter 4 is joint work with Francisco Blasques and Artem Duplinskiy and builds on the fact that university students consistently evaluate their female teachers lower than male teachers. The gender gap in evaluations is robust to controlling for such proxies of teaching effectiveness as student grades and self-study hours (Mengel et al., 2019). However, observational data do not allow us to infer whether this discrepancy stems from a behavioral bias on the side of the students, or there are unobservable characteristics of teachers that account for the difference in evaluations.

We designed an experiment that exogenously varies the perceived gender of the teacher without deceiving the subjects. We aim to recruit male and female teachers to make a videolecture on a mathematical topic; the lecture takes the form of a short video with self-introduction and narrated slides. The subjects in the laboratory need to watch the lecture, take an incentivized quiz on its content and subsequently evaluate their instructor. This design minimizes verbal and non-verbal differences in teaching style, allowing us to make causal inference about the presence of gender bias in student evaluations. On top of that, we will present two versions of the lecture content: one of high quality and another one of intentionally degenerated quality. This attempt to exogenously manipulate the quality of teaching will be the first to shed the light on the response of gender bias in evaluations to variations in the true effectiveness of the teacher.

Finally, we propose a transparent communication of teachers' incentives to the students as an intervention to overcome potential gender bias in evaluations (if any). An average undergraduate student is unlikely to be aware of the incentive structure in academia, including the fact that his evaluation might affect the teacher's future career. We hypothesise that communicating these incentives might make students evaluate their teachers more carefully, potentially making the evaluations of the female teachers less emotional and more objective. In our experiment, the subjects in the role of the teachers compete for a monetary prize given to the teacher with the highest evaluation scores in our experiment. Depending on the treatment, this prize will either be not mentioned or explicitly stated to the students before they proceed to the evaluations. This kind of intervention is much less suggestive than the previous attempts to ask the students to avoid gender discrimination.

Chapters 1 and 4 of this dissertation are self-contained and can be read independently. Chapters 2 and 3 represent the theoretical and the empirical parts of one study. Every chapter has its own appendix. All appendices and the bibliography can be found at the end of the dissertation.

1 Two Heads are Better than One: Teams and Individuals in Standard Auction Formats

1.1 Introduction

Auctions are a very important allocation mechanism. In the theoretical and empirical literature, bidders in auctions are usually assumed to be (representative) individuals. However, often, small groups or teams (henceforth used interchangeably) determine bids or bidding strategies together (see, e.g., Börgers and Dustmann, 2005). Teams may be organized formally, potentially including a hierarchy, or informally such as it is usually the case with families or teams of experts without a formal structure.

Examples of bidding in teams readily come to mind. Think of a family that decides about their bid in an online auction. However, also on the level of corporate auctions, teams play a crucial role. In spectrum auctions (Abbink et al., 2005) or auctions for offshore oil leases, teams of managers or experts (geologists) formulate strategies for bidding (Capen et al., 1971, Hoffman et al., 1991). Interestingly, we know very little about the extent to which existing knowledge from bidding behavior of individuals can be transferred to auctions that involve team bidders. One reason is that the existing evidence, mostly from laboratory experiments, is not fully conclusive. We discuss the relevant literature in greater detail below.

This paper, therefore, compares individual and team bidding in a systematic design. We implement three commonly used auction formats: a first-price auction, a second-price auction, and an ascending English auction. Using a laboratory experiment, we systematically vary the type of decision maker by having either individuals or small unitary teams deciding about the bids. Our focus is on unitary teams, i.e. teams that do not face an internal conflict in terms of monetary outcomes and that have to come up with a joint decision after some deliberation. Obviously, team decision settings are quite diverse outside the laboratory. However, unitary teams seem to provide a good first impression of potential differences between individual decision makers and teams, and more involved teams settings follow immediately as potential extensions.

It is well-known that individuals and small unitary teams differ in several respects (Charness and Sutter, 2012, Kugler et al., 2012). First, individuals and teams differ in their ability to avoid mistakes (Bornstein et al., 2004, Blinder and Morgan, 2005, Cooper and Kagel, 2005, Sutter, 2005, Gillet et al., 2009, Feri et al., 2010) or in their learning ability and pace (Kocher and Sutter, 2005, Kocher et al., 2006, Maciejovsky et al., 2013). According to the economists' definition, teams are thus usually more rational decision makers than individuals. Applying results from existing studies using different decision making games would lead to the expectation of fewer instances of overbidding in first-price auctions and of fewer deviations from the dominant strategy of bidding one's own valuation in second-price auctions.

Second, individuals and teams might potentially differ in the risks they take. There is a long discussion starting in the early psychological literature about whether there is a "risky shift" from individuals to teams (Stoner, 1961) or whether the shift is actually "cautious" (Moscovici and Zavalloni, 1969). Recent evidence from controlled laboratory experiments in economics provide inconclusive results (Baker et al., 2008, Shupp and Williams, 2008, Masclet et al., 2009, Zhang and Casari, 2012). Obviously, a difference in revealed risk attitudes from individuals and teams would affect equilibrium bids in first-price auctions, not in second-price auctions.

Third, there could be a difference in the emotional reactions to events during an auction or between auctions.[2] Auction fever describes a non-rational immersion into an auction that might lead to overbidding. There is large literature in social psychology showing that team decision making might go out of hand and lead to even disastrous consequences. This literature that is often summarized under the heading "groupthink" (Janis, 1972, 1982, 1989) has looked at (often spectacular) misjudgments or failures that followed deliberation within teams. Usually, groupthink goes hand in hand with some deficiencies in the group decision making process such as the existence of a strong group leader together with the suppression of deviant opinions. While there is no evidence yet, it is conceivable that groupthink could be related to emotional reactions or over-competitive attitudes in auctions. Does auction fever perhaps affect individuals and teams differently? Potential differential effects of auctions fever should be observable in ascending English auctions.

Our list of potential influences on bidding in auctions that lead to differences between individual bids and team bids is not exhaustive. There might be additional relevant aspects, but those mentioned above are probably the most important ones. So

[2]See, for instance, Breaban and Noussair (2018) for an application of emotion measurement in experimental asset markets.

far, the existing experimental literature on bidding in auctions has almost exclusively focused on individual bidders. To the best of our knowledge, there are only a few exceptions. Cox and Hayne (2006), the first paper on the difference between individuals and teams in auctions, focus on the effect of information on team bidding and less on a controlled comparison between individual and team bidders. They find that more information in a common value auction environment makes teams less rational bidders than individuals – an effect that they call "curse of information". Indeed, such a curse could be related to groupthink or over-competitiveness. Interestingly, Sutter et al. (2009) find a similar result in a different framework. Their focus is on an ascending sealed-bid English auction with a private and a common value component (with some resemblance to spectrum auction designs). They show that teams stay significantly longer in the auction and that they also pay higher prices. As a consequence, they make smaller profits and fall prey to the winner's curse more often.[3]

We are not aware of any study on the differences between individual and team bids in the most straightforward auction designs: sealed-bid first- and second-price auctions. We also implement an ascending English auction. An important feature of our design is that we keep all aspects equal across the individual treatments and the team treatments – hence, the bids in the auctions formats can be compared directly. Our teams consist of three members that remain anonymous and can only interact through a real-time chat in order to reach a joint decision. We elicit cognitive abilities and risk attitudes on the individual and on the team levels, alongside a set of other background measures. We also analyze the emotions of our decision makers as an additional control. Individuals always bid against individuals on a market consisting of three bidders, and team always bid against teams on a market consisting of three bidder teams, and this is common knowledge among participants of the experiment.

We find that individuals are more prone to overbidding in first-price auctions and less likely to bid their own valuation in second-price auctions. This is in line with our first expectation that teams are more rational decision makers than individuals. Indeed, we can show that the significant treatment effects vanish, as soon as we control for both cognitive ability and risk attitudes, in line with our second expectation. Taken together,

[3]Casari et al. (2016) look at a company takeover experiment that reports less overbidding by teams than by individuals. Similarly, Sheremeta and Zhang (2010) study Tullock-contests, finding again less overbidding by teams.

in these simple formats, teams are the better decision makers in maximizing profits, which is in contrast to the existing results for more complex formats (Cox and Hayne, 2006, Sutter et al., 2009). Our results provide a rationale for reconciliation of the existing evidence for team decision making in auctions with the general picture that is painted regarding team decision making. Teams are more rational decision makers, but there might be forces such as increased complexity or strongly competitive environments in which teams fare worse than individuals, potentially due to inherent problems of teams as decision makers such as groupthink. Interestingly, we do not find any difference between individuals and teams in the ascending English auction. However, our design is such that it was slightly tilted against auction fever and groupthink, because we use a private value setup and a relatively slow-paced decision making environment.

The rest of the paper is organized as follows. In Section 1.2, we describe the details of the experimental design. Section 1.3 reports the results of our experiment. Section 1.4 discusses our findings, and Section 1.5 concludes the paper. Regression tables and instructions for the experiment can be found in Appendix A.

1.2 Experimental design

1.2.1 General structure and treatment variation

The experiment consisted of four parts. In the first part subjects participated in a series of auctions (either first-price sealed-bid auctions or English auctions). In the second part they bid in one round of a second-price sealed-bid auction. In the third part of the experiment subjects completed a risk elicitation task, where they had to make an investment into a risky asset. It was followed by a Raven's matrices test in the fourth part to measure cognitive ability. Finally, we concluded the experiment with a short questionnaire.

We varied our treatments in two aspects in a 2×2 factorial design. The first treatment difference concerned the type of the auction in the first part of the experiment. In part 1 of the *FPA* treatments, subjects participated in repeated first-price sealed-bid auctions. In part 1 of the *EA* treatments, participants bid in repeated English auctions. The detailes of the rules of both auction formats are explained in detail below.

The second treatment variation was the type of the decision-making entity. In *IND* treatments, all subjects made their choices individually. In *TEAM* treatments, subjects made decisions in teams of three. These teams stayed the same throughout all four parts of the experiment, but subjects did not know the identity of the other team members. The decision making process within teams was the same throughout all parts of the experiment: first, all three members of a team made an individual proposal for their team's bid in an auction (or an investment in part 3, or a test answer in the Raven's test in part 4). Afterward, if the individual proposals did not coincide within a team, its members had to agree on a joint decision within a limited time. During this time, team members could discuss their choices in a real-time chat. Once an agreement was reached within each team, the experiment proceeded. If some team failed to agree on a joint decision within the specified time frame, its members did not earn anything in the respective auction (or the respective task in part 3 or part 4). To make the treatments more comparable in terms of deliberation, we asked subjects in *IND* treatments to write a short explanation of their choice after they had made it. While writing the explanation, individuals could also change their decision if they wanted so.

1.2.2 The rules of the experiment

Part 1: repeated first-price and English auctions. In the first part of the experiment, bidders participated in a finite number of repeated auctions. We use the term "bidder" with respect to both individual bidders in *IND* treatments and team bidders consisting of three subjects in *TEAM* treatments. In the beginning of each auction, bidders were randomly matched into markets of three to bid for a virtual good. Bidders had independent private values for the good, drawn from a uniform distribution on $[0, 100]$ experimental currency points.[4] After learning their private values, bidders submitted their bids for the good. As described above, the bid submission process corresponded to individual bid proposals followed by a discussion in *TEAM* treatments, and an individual bid choice followed by an explanation in *IND* treatments. In the following, different auction rules in the *FPA* and *EA* treatments applied:

(1) In *FPA* treatments, bidders submitted their bids only once per auction. After all three bidders on a market submitted a bid, the auction ended. The bidder with the highest bid on the market won the auction and paid his own bid for the good.

[4]The members of the same team in *TEAM* treatments had the same values for the good so that teams did not have more information than individual bidders in *IND* treatments.

The payoff for the winner was thus the difference between her own value and her bid. For teams it means that each member of a team received the same payoff equal to the difference between the team's value and the bid, thereby facing exactly the same monetary incentives as bidders in *IND* treatments. Bidders who did not win the auction got 0 points.

(2) In *EA* treatments, each auction consisted of several bidding rounds. In each round bidders could either submit a bid higher or equal to the minimum bid in the respective round or quit the auction (equivalent to submitting a bid of zero). The minimum bid in each round amounted to the highest bid on the market in the previous round, referred to as current price, plus an increment. The increment decreased with the current price level as laid out in Table 1.1. The auction ended when only one bidder was left on the market. This bidder won the auction and received a payoff equivalent to the difference between her valuation and her own bid from the last round. In *TEAM* treatments this payoff was received by each member of the winning team. As for the first-price auction, bidders who did not win the auction got 0 points.

Table 1.1: Increment rule in *EA* treatments

Conditions	Increment	Minimum bid
current price ≤ 30	10	current price + 10
$30 <$ current price ≤ 45	5	current price + 5
$45 <$ current price ≤ 90	3	current price + 3
current price > 90	2	current price + 2

When an auction ended, the participants received feedback about winning or not winning the auction and the price at which the good was sold. Then, the next auction started with exactly the same rules as before. The markets were randomly re-matched, and the new valuations were drawn (note that in *TEAM* treatments the composition of the teams remained the same). There were 12 auctions in *FPA* treatments and 8 auctions in *EA* treatments.[5] Only one auction was selected as relevant for payment at the end of the entire experiment. In the instructions we explicitly hinted at the

[5]We conducted a different number of auctions in *FPA* and *EA* treatments merely due to time considerations. English auctions on average took longer than first-price auctions.

possibility of making losses if the winning bid exceeds the winner's private value. The highest possible bid in all treatments was 110 points.

Part 2: second-price auction. After finishing the first part, subjects participated in a one-shot second-price sealed-bid auction. Since the optimal bidding strategy in the second-price auction is relatively straightforward and boils down to bidding own private value, we consider bidding behavior in this part of the experiment as a potential proxy for a bidder's rationality.

Like in the first part of the experiment, bidders were matched into markets of three to bid for a virtual good. In *TEAM* treatments subjects were bidding in the same teams as before. The values were private, independent and uniformly distributed over $[0, 100]$. After all bidders submitted their bids, the auction ended. The bidder with the highest bid on the market won the auction and received her private value net of the second highest bid on the market. To avoid potential income effects, we gave the participants feedback about the outcome of part 2 only at the end of the experiment. We also explicitly hinted at the possibility of making losses if the second highest bid on the market exceeds the winner's private value.

Part 3: risk elicitation task. In this part of the experiment subjects made a decision about an investment into a risky asset, based on the approach of Gneezy and Potters (1997). As before, in *IND* treatments participants made their choices individually, while in *TEAM* treatments they agreed on a common investment within the teams.

Each decision-maker - individual or team - received an endowment of 100 experimental currency points and could invest any amount $R \in [0, 100]$ into a risky asset and keep the rest $100 - R$. The return from the investment was either $2.5R$ or 0 with equal probabilities. When the decision makers chose their investment R, the outcome was randomized and participants learned how much they earned. As before, subjects in teams shared the same payoff, having exactly the same incentives as subjects in *IND* treatments.

Part 4: Raven's matrices test. The last part of the experiment was a short version of Raven's progressive matrices test, aimed at obtaining a measure of decision makers' reasoning abilities. The test consisted of eight tasks to identify a symbol that fits the given pattern (see Appendix A.2). The time for solving each problem was limited to 90 seconds. If a decision maker managed to solve a problem correctly within this time frame, she got 50 euro-cents for the solution (also implying 50 cents per team member in *TEAM* treatments). If she submitted no answer or the answer was wrong, she did not earn anything for the respective problem. Once the subjects went through all eight problems, they learned how many of their answers were correct and how much they earned.

Finally, the participants filled out a short questionnaire and received feedback about their earnings from the four parts of the experiment.

1.2.3 Laboratory protocol

The experiment was conducted in the experimental laboratory MELESSA at the University of Munich, Germany, using the experimental software zTree (Fischbacher, 2007) and the organizational software ORSEE (Greiner, 2015). A total of 252 subjects, mostly undergraduate students, participated in 14 sessions with 18 subjects each. Sessions in *FPA* treatments lasted about 90 minutes, sessions in *EA* treatments took about two hours. Average earnings in the experiment amounted to 21.10 euros, including the show-up fee of 4 euros. Subjects received written instructions (see Appendix A.2) for each part of the experiment, after the previous part was finished. They knew that there would be four parts from the beginning. We videotaped the subjects during the experiment and used FaceReader emotion recognition software. All participants had to sign an extra consent form for the video; there were no dropouts.

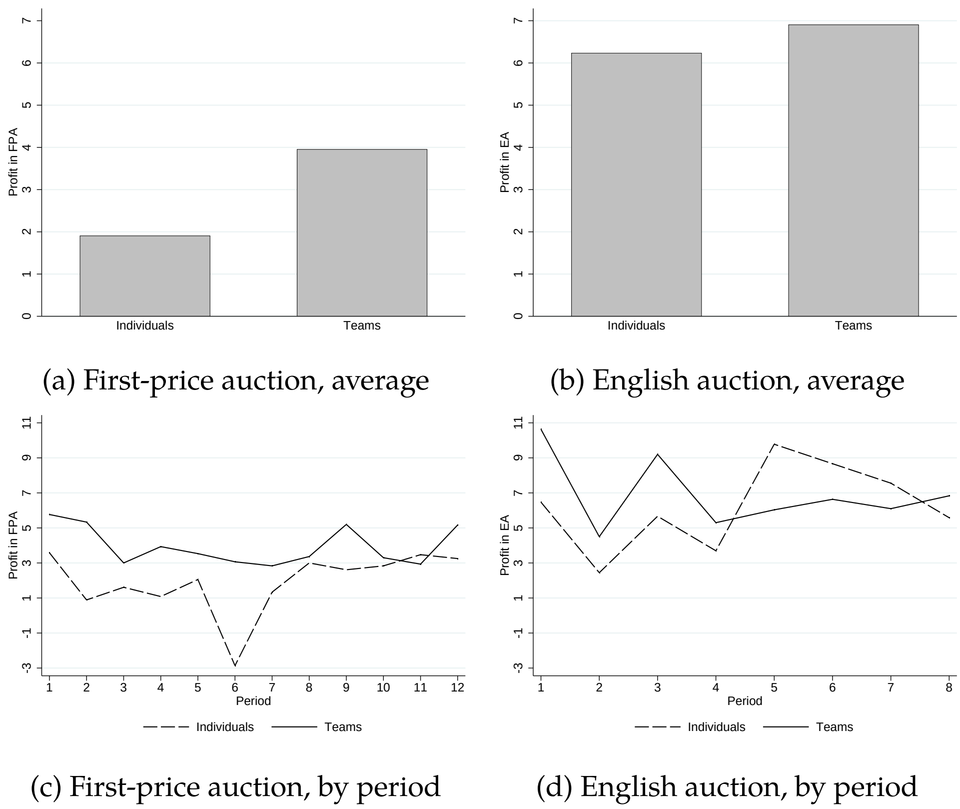
Since in *TEAM* treatments subjects were matched into teams of three, in part 1 and 2 there were two markets with three team bidders each. In each round, six team bidders in one session were randomly rematched into two markets, leaving us with one independent observation per session. In *IND* treatment sessions there were six markets with three individual bidders. We reshuffled the markets after each period within matching groups of six to keep the probability of facing the same people on a market equal across treatments. Thus, in *IND* treatments we had three independent observations per session. The subjects were aware of the matching protocol. We conducted two sessions each in *IND FPA* and *IND EA* treatments, and five sessions each in *TEAM FPA* and *TEAM EA* treatments.

1.3 Results

In this section we analyze the profits earned by bidders, different characteristics of bidding behavior specific to first-price and English auctions, and market outcomes such as prices and the efficiency of the auctions.

1.3.1 Realized profits

We start with analyzing actual profit that bidders earned in the auctions. By design, winner's profits in the first-price auction amount to the difference between their private values and their bids. If a bidder did not win an auction, the profit was zero. Our data show that in first-price auctions teams earned on average twice as much as individuals: 3.95 and 1.91 experimental currency points, respectively ($p = 0.0446$; two-sided Wilcoxon rank sum test; $N = 11$). Figures 1.1a and 1.1c show the average profit in first-price auctions and its dynamics over the twelve periods for individuals and teams. What could be a possible explanation for this difference? We attempt to answer this question by analyzing the effect of team membership on profits in first-price auctions using panel regressions with random effects, with robust standard errors clustered on



(a) First-price auction, average (b) English auction, average

(c) First-price auction, by period (d) English auction, by period

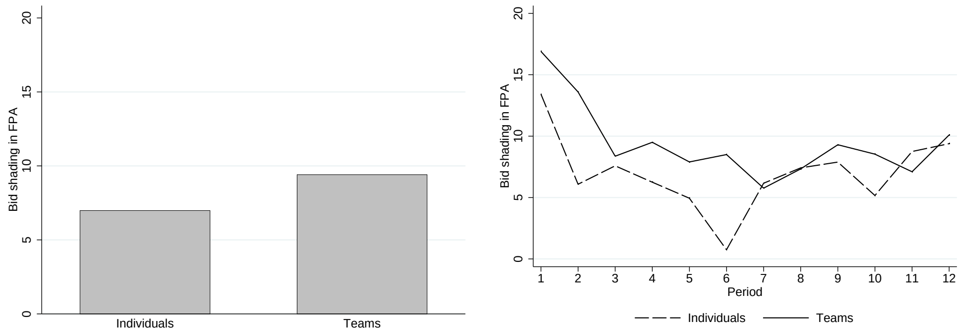
Figure 1.1: Profits earned by bidders

the matching group level. We use the bidding behavior in the second-price auction, the amount invested into the risky asset, the Raven's test score and several demographic characteristics as control variables. It can be seen in Table A.1 in Appendix A.1 that being an individual bidder reduces the profit by more than 2 points on average compared to team bidders. However, profit increases over time for individuals, implying that individual bidders might learn and start to close the gap. Profit is highly correlated with the performance in the second-price auction: if an average bidder bids closer to the optimal strategy in the second-price auction, she earns significantly more in the first-price auction as well. Bear in mind that the second-price auction takes place after experiencing the first-price auctions; hence, causality cannot be established for this statement. There is a weakly significant correlation of profits with risk attitudes (relatively less risk averse subjects earn more) and, interestingly, gender also has a strong effect on profits: female participants on average earned 2.6 points less then male participants, controlling for everything else.

In contrast, in the English auction we find no difference between the profits earned by individuals and teams. An average individual bidder in the English auction earned 6.23 points compared to 6.90 points in the *TEAM EA* treatment. This difference is clearly not statistically significant (see Figures 1.1b and 1.1d). Regression analyses in Table A.4 in Appendix A.1 also show that there is no significant treatment effect on profits in our English auctions. However, like in first-price auctions, there is weak evidence of individual learning and a strong correlation between profits earned in the English auction with bidding behavior in the second-price auction. In addition, profit is positively correlated with the self-reported high school math grade.

1.3.2 Bidding behavior

Bid shading is an important characteristic of bidding behavior in first-price auctions. We define bid shading in *FPA* treatments as the difference between the private value and the submitted bid. Thus, for the winners of the auction, bid shading is equivalent to the realized profit, while it also provides information about the behavior of the bidders who did not win the auction. Figure 1.2a shows that individuals on average bid 6.98 points below their valuation, and teams shade their bids more – by 9.41 points. Using the most conservative test – a two-sided Wilcoxon rank sum test on the level of independent observations – the difference is not significant due to the small number of strictly statistically independent observations.

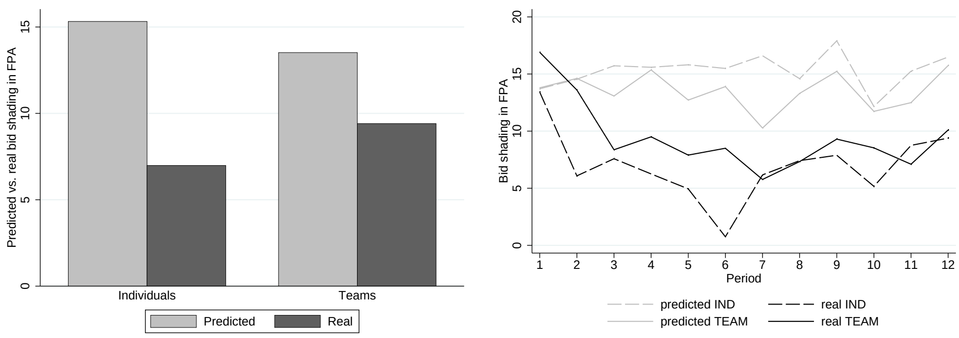


(a) Average (b) By period

Figure 1.2: Bid shading in first-price auctions

This changes when we look at parametric results. We estimate regression models of bid shading in Table A.2, where the actual bid in the first-price auction is the dependent variable, regressed on the private value of the good. The coefficient associated with the value of the good can be interpreted as bid shading, which turns out to be slightly below 18% of the value in all specifications. The regression specifications (1)-(3) show that there is a difference between bidding behavior of individuals and teams, that vanishes when we control for variables from other parts of the experiment in specifications (4)-(8). In line with the results from subsection 1.3.1, bid shading is negatively correlated with the deviation from rational behavior in the second-price auction. Moreover, bidders with higher Raven's test scores and better math grades submit lower bids, suggesting that general intelligence matters for bidding strategy. Subjects with longer past experience of participating in economic experiments also submitted lower bids in the first place. Risk aversion is associated with slightly higher bids: more risk-tolerant bidders submit lower bids, thereby increasing potential profit, but lowering the chance of winning. Finally, specifications (6)-(8) reveal that bids increase over time in *TEAM* treatments, suggesting that teams might become more competitive over time. Figure 1.2b shows that until period 7, average bid shading was higher in teams, but from period 7 on, there is little difference.

We have seen how bid shading differs across treatments, but how far away are the actual bids from the optimal strategy in the first-price auctions? To answer this question, we predict optimal bids using a simple parametric model with risk averse bidders. In this model, we assume that bidders have a constant relative risk aversion utility function $u(x) = x^{1-r}/(1-r)$, where x is the monetary payoff and r is the risk aversion parameter ranging from 0 to 1 (a higher r corresponds to higher levels of risk aversion). We get a proxy for this parameter from the choices in the risk elicitation task in part 3 of the experiment. The estimated average levels of risk aversion are $r = .07321$ for individuals and $r = .1721$ for teams: teams were more risk averse than individuals, but the difference is not significant.[6] Using these estimates and the realized values for the good, we predict optimal bid shading (value minus bid) for every bidder in every single auction. The optimal bid shading is presented together with the actual bid shading in Figure 1.3. We observe that both individuals and teams shade their valuations less, i.e. overbid compared to the prediction. This is a common finding in the auction literature. However,the deviation from the rational bidding strategy is more pronounced for the individuals ($p = 0.1003$; two-sided Wilcoxon rank sum test; $N = 11$). A regression with individual bid shading as the dependent variable gives very similar results as those in Table A.2. Again, bidding in the second-price auction and Raven's test scores matter.



(a) Average (b) By period

Figure 1.3: Deviation from the estimated optimal bid shading in first-price auctions

In contrast to the *FPA* treatments, we find no difference in bidding behavior of individuals and teams in English auctions. One characteristic of bidding behavior specific to English auctions is the average number of rounds that bidders spend in an auction before they drop out. We observe that teams stay slightly longer in English auctions: team bidders were on average active in 4.06 rounds per auction compared

[6]In fact, this difference was purely due to a higher number of teams that chose to invest zero points into the risky asset. We assumed $r = 1$ for those who did so, creating discontinuity in the parameter. Four out of 30 teams invested nothing in this task, compared to one out of 36 individuals. If we exclude these observations, the difference in risk preferences goes away.

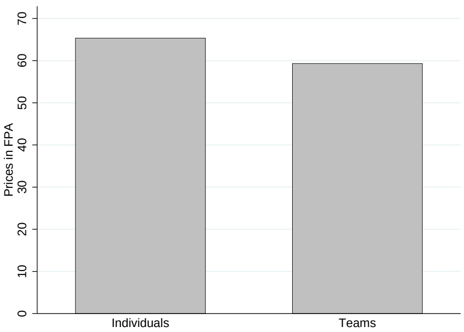
to 3.27 rounds for individual bidders. However, this difference is not significant on conventional levels. Another characteristic of bidding behavior in the *EA* format is the relative bid in each round. We define it as the ratio of the actual bid to the minimum bid in the respective round. The average relative bids were not significantly different: they are 1.48 for individuals versus 1.29 for teams.

1.3.3 Market outcomes: prices and efficiency

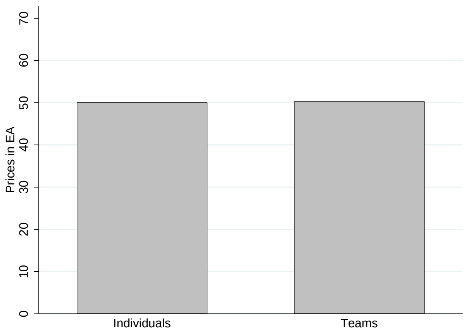
Given that the valuations were assigned randomly and independently, lower profits observed among the individuals in *FPA* treatments implies that individuals paid higher prices than teams. Indeed, in first-price auctions the average price for the good amounted to 65.34 points in *IND* treatments compared to 59.32 points in *TEAM* treatments ($p = 0.0679$; two-sided Wilcoxon rank sum test; $N = 11$). Average prices and their dynamics are depicted in Figures 1.4a and 1.4c. The regression analysis (Table A.3) lacks power to capture the treatment effect since the data have to be collapsed on the market level; however, we still find the correlations between prices and control variables. In line with the observed bidding behavior described above, prices are positively correlated with the deviation from rational bidding in the second-price auction and risk aversion, and negatively correlated with the score in the Raven's test, high school math grade and previous experience of participating in experiments.

In English auctions, individuals and teams paid almost exactly the same price, slightly above 50 points (see Figure 1.4b). Although there is no difference on average over all periods, the development of the prices was different in *IND* and *TEAM* treatments. As one can see in specifications (5)-(7) in Table A.5, prices paid by teams increased over time, while prices paid by individuals decreased. This pattern is also reflected in Figure 1.4d.

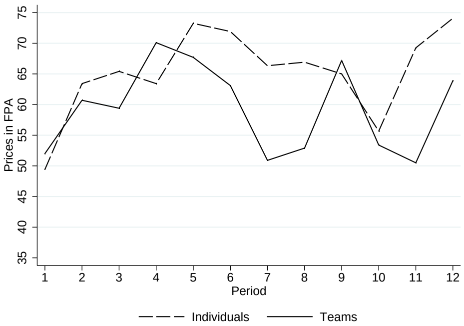
Finally, we look at efficiency, which is an important variable for auction design. An auction is defined as efficient if the bidder with the highest private valuation obtains the good. In *FPA* treatments with individual bidders 78% of all auctions were efficient, while in *TEAM FPA* treatments this share was 89%, the difference being not significant on conventional levels. Efficiency in English auctions was almost the same in the two treatments and almost the same as in the *IND FPA* treatment – slightly above 78%.



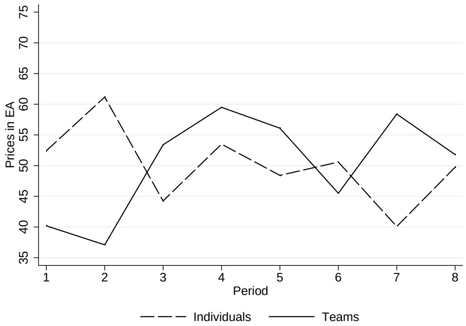
(a) First-price auction, average



(b) English auction, average



(c) First-price auction, by period



(d) English auction, by period

Figure 1.4: Realized prices

1.3.4 Emotions

Our experiment was recorded on camera and the facial expressions of the subjects were analyzed using the facial recognition software FaceReader. We focus here on investigating the role of emotions in *FPA* treatments because of the relatively simple structure of first-price auctions. In particular, we were interested in the emotions experienced when the subjects saw their feedback screens at the end of each period, on which they learned whether they won the current auction or not and how much they earned. We expected the expressed emotions to have an effect on the future decisions of the bidders in subsequent auctions.

Apart from a neutral facial expression, the software detects six basic emotions: anger, happiness, surprise, disgust, sadness, and fear. Their magnitude was measured as a number between 0 and 1, captured multiple times per second. For each subject in each period, we took an average over those measurements within the first ten seconds that the subjects saw their feedback screen and used these measures for our analysis (synchronization between zTree and FaceReader was achieved by employing the software μCap by Doyle and Schindler, 2019). An additional composite variable "valence" amounts to the difference between measured happiness and the strongest negative emotion (anger, disgust, sadness, and fear). It reflects how positive or negative the overall emotional state is and ranges from –1 to 1.

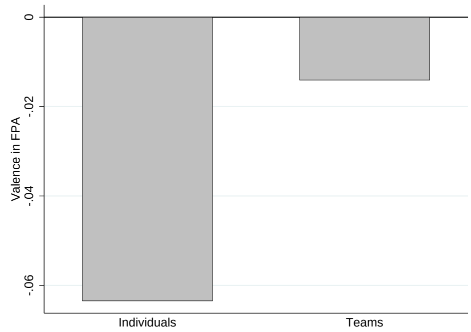


Figure 1.5: Valence of the emotions experienced by the subjects seeing their feedback screens in *FPA* treatments

Our first interesting observation suggests that subjects in *TEAM* treatments experienced on average more positive emotions than subjects in *IND* treatments (see Figure 1.5; $p = 0.1003$; two-sided Wilcoxon rank sum test; $N = 11$). This could possibly be explained by the fact that teams earned more on average and thus were more positive about the outcomes of the auctions. Another potential reason could be an inherent preference for interaction in teams (Kocher et al., 2006). Our *TEAM* treatments were likely to be more interesting for subjects due to the interaction through the chat messages and the challenge of agreeing on a joint bid. This experience could have influenced the overall satisfaction with the experiment.

To test whether the measured emotions reflect plausible reactions to the outcome of an auction, we checked whether different emotional states are correlated with winning or losing an auction in the *current* period. Table A.6 shows that winning an auction is associated with more happiness and less anger, and respectively, a greater valence. Thus, we may conclude that facial recognition of emotions worked properly in our experiment. Interestingly, individuals seem to experience a bit more sadness than team members when not winning the auction.

The most interesting question lies in the potential impact of emotions on *future* bidding behavior. We evaluate this effect by estimating our regression models of bid shading with emotions experienced in the previous period as control variables. The results are reported in Table A.7. We observe that two emotions are significantly correlated with bidding behavior in the next period. First, if subjects experienced sadness, it had a negative impact on their next bid, i.e. they bid, on average, closer to the rational prediction. One channel could be that winners were sad about the small profit they made despite winning the auction; in the next round, they would adjust their bids downwards to increase the expected profit. Second, the feeling of fear was positively correlated with the bids in the next period. If a subject was afraid of losing the next auction and ending up with zero profit, she might have overbid even more to increase the chances of winning at least something. This logic is similar to risk aversion, supporting the literature that claims an association between fear and risk-taking (Nguyen and Noussair, 2014). While the effect of sadness would become weakly significant when controlling for the number of emotions covered (multiple hypotheses testing), the effect of fear is robust. Interestingly, sadness and fear are emotions that one would not associate with auction fever. Bear in mind, however, that a seal-bid first-price auction is *ex ante* much less prone to auction fever than a fast-paced ascending English auction or a descending Dutch auction.

1.4 Further analyses and discussion

The results of the experiment have shown that teams bid closer to the rational prediction in first-price auctions, paying lower prices and earning larger profits than individuals, but we observe no difference between team and individual bidding behavior in ascending English auctions. Why do teams perform better in only one of the auction types? Although both auctions in our experiment were implemented in the simplest setting with independent private values, there seems to be a difference in the complexity of the optimal bidding strategies. In the English auction, the dominant strategy is rather straightforward: a bidder should stay in the auction, as long as the minimum bid does not exceed her private value, and drop out otherwise. In first-price auctions, however, a bidder needs to figure out the best level of bid shading, taking into account the distribution of the values, her own risk attitude, and risk attitudes of the competitors on the market.

We see that teams outperform individuals in this relatively complex decision-making environment. Our analysis suggests that this effect can be explained by differences in reasoning abilities of teams and individuals: the performance in first-price

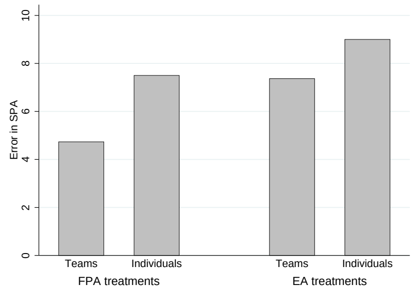
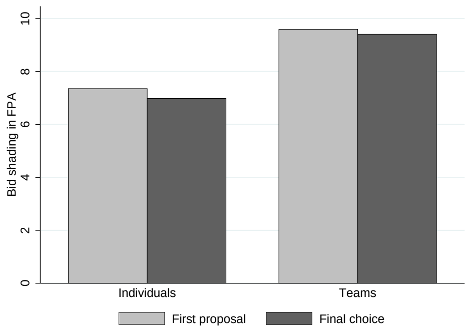


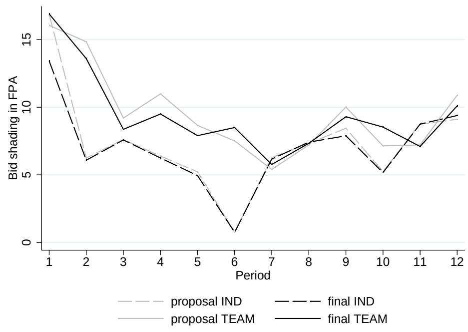
Figure 1.6: Absolute deviation from rational strategy (value minus bid) in SPA

auctions is correlated with the choices in a second-price auction, i.e. smaller deviations from the dominant strategy in the latter lead to less overbidding in the former. We consider the choices in the second-prices auction a measure of reasoning ability, but we get similar results for the scores in a Raven's test, the high school math grade, and the previous exposure to laboratory economic experiments. Among those, the deviation from rational bidding in the second-price auction is the most robust predictor of overbidding in first price auction. Indeed, average second-price auction bids in *TEAM* treatments are closer to the values than in *IND* treatments, see Figure 1.6. Non-parametric tests lack power to render this difference significant, but the discrepancy in this proxy for rationality can predict the treatment difference in *FPA* treatments, while it does not play a role in the relatively simple decision-making environment of *EA* treatments.

One could ask whether the observed difference in first-price auctions could be partly attributed to the fact that team members adapt their choices when they agree on a common strategy. In that case we would see a difference between final bids and initial bid proposals that were submitted individually before entering the discussion in *TEAM* treatments. Figure 1.7 depicts initially proposed bid shading along with actual bid shading (as mentioned in section 2, individual bidders also had the possibility to change their first choices). We see that the initial proposals are not different from the final choice of both individuals and teams, suggesting that communication and agreeing on a common strategy cannot fully account for the observed difference in behavior. There seems to be a component in team decision making that makes individuals change their revealed preferred decision already prior to the team interaction. Perhaps anticipated peer observation makes them think harder about the optimal strategy in first-price auctions. It could also be the case that team members agreed on a common strategy early on and adjusted their future bid proposals accordingly. The latter speculation is



(a) Average



(b) By period

Figure 1.7: Proposed and final bid shading in first-price auctions

supported by the observation captured in Figure 1.7b: the initial proposals of individual and team bidders were very close in the first period, but followed different patterns from period 2 on.

Since auction experiments are inherently complex, we want to make sure that our results are not driven by lack of understanding of the experimental setup. To tackle this issue, we check whether there are bidders that consistently bid above their valuations, despite a clear statement in the instructions that losses are possible in this case. Remarkably, in both *TEAM* treatments we do not observe a single bid above the private values. Only in the *IND FPA* treatment there are two subjects who always overbid and do not learn throughout the course of the experiment. If we exclude those observations from the analysis, it does not qualitatively change our results. However, the fact that we observe these outliers only in *IND FPA* treatment is interesting itself, supporting our claim that individual bidders have a harder time to cope with complex auction environments.

Apart from observing the realized profits and prices, unfortunately, there is not much more to say about the bidding behavior in *EA* treatments. Typically, in a setting with independent private values, both relative bids and the number of active rounds could be considered as a proxy for rational behavior. The more subjects bid relative to the minimum possible bid, and the longer they stay in an auction, the more competitive and the less rational their behavior often is. We observe no significant difference between teams and individuals with respect to relative bids and the number of active rounds, but in our experiment these potential measures of rationality turned out to be very noisy. Our *EA* treatments lasted about two hours, on average, and subjects might have become a bit tired during the experiment, making this clear in the discussion stage. Thus, there was a potentially perfectly rational motivation for both teams and individuals to bid higher than the minimum bid and drop out faster in order to finish the experiment more quickly. This property of our experimental setup might partly account for the absence of evidence for auction fever on the side of the teams, contrary to the previous findings.

If a seller had the possibility to choose between the two auction types in our setting, first-price auctions would be more profitable than English auctions, and vice versa for the bidders. The revenue equivalence principle holds neither in *IND* nor in *TEAM* treatments, and realized prices in first-price auction are significantly higher than in English auctions, as once can see on Figure 1.4 ($p = 0.0014$; two-sided Wilcoxon rank sum test; $N = 22$). This is partly explained by bidders' risk aversion that makes them bid higher in first-price auctions. However, as mentioned above, both individuals and teams overbid in comparison to the theoretical prediction that takes risk aversion into account, boosting the difference in profitability between *FPA* and *EA* treatments.

Finally, some specifications of the regression models allow us to draw inferences on the dynamics of bidding behavior and auction outcomes. On the one hand, in both auction types we find that the performance of individuals improves over time: for instance, individual profits increase in both first-price and English auctions. We interpret this as learning, which is more pronounced among individuals, since their average performance is poorer, than in teams and, therefore, there is more room for improving the bidding strategy over time. On the other hand, in contrast to this pattern, the bidding behavior of teams became less rational over the course of the experiment: team bids in first-price auction and prices in English auction increased with the time. This deterioration in performance supports the strand of the literature that shows that teams tend to behave more competitively than individuals (as in Cox and Hayne, 2006, Sutter et al., 2009).

1.5 Conclusion

We have conducted an experimental study that systematically compares the bidding behavior of individuals and unitary teams in different auction formats. Our results show that individuals deviate more from the optimal strategy in first-price and second-price auctions than teams, while there is no difference between individual and team bidders in English auctions. The difference in overbidding in first-price auctions between individuals and teams is strongly correlated with bidders' ability to make rational decisions, which we consider to be reflected in their bids in second-price auctions and in their Raven's cognitive ability test scores. In other words, teams outperform individuals in first-price auctions, because they are more rational decision makers. This finding challenges the results from the existing literature that showed that teams perform worse than individuals in auctions (Cox and Hayne, 2006, Sutter et al., 2009) and it is in line with the studies where teams outperform individuals in other decision-making environments (e.g., Casari et al., 2016, Sheremeta and Zhang, 2010, Sutter, 2005). We argue that the complexity of the decision making environment matters for the observed difference in individual and team behavior: since the optimal bidding strategy in first-price auction is more complex relative to our English auction format, better reasoning abilities of the teams play a role in determining the difference in first-price auctions, but not in our simple version of the English auction. Also contrary to our expectations, we find no evidence for auction fever among teams in English auction, with this result possibly being an artifact of our experimental setup with independent private values and without time pressure.

Our paper is the first that systematically analyzes different auction formats with individual and team bidders. Obviously, there are choices to make, when it comes to the experimental design, and thus there is room and need for more research. First, one could investigate whether more systematically varying the complexity of an auction affects the performance of individuals and teams in a different manner. In our setting we found that teams outperformed individuals in more complex auctions; however, the previous studies that observed teams performing worse than individuals had even more complicated auction rules. How team and individual bidders respond to exogenous variations in complexity of an auction, keeping everything else constant, and whether the difference in bidding behavior is monotone in complexity, remains an open question. Second, it would be interesting to learn about which auction features could potentially lead to groupthink and over-competitiveness among teams, which were documented in some of the previous studies. Our setup was clearly tilted away from groupthink phenomena. Introducing some (natural) hierarchy could be an interesting avenue for future research. Our experiment also shows that an ascending price is not enough for teams to exhibiting anything close to auction fever. Finding the characteristics of an auction (potentially time pressure, face-to-face interaction, etc.) that trigger those reactions in teams and, hence, potentially deteriorate their performance, would be an important aspect for auction design that we think requires more attention.

2 Optimal third-degree price discrimination and reciprocity: A theoretical model

2.1 Introduction

Third-degree price discrimination, or charging different prices for the same product to different consumers, is one of the tools for firms to increase their profits. According to classical industrial organization, price discrimination is always more profitable than uniform pricing (see Armstrong, 2006). Third-degree price discrimination has been observed on different markets, including markets for cars (Verboven, 1996), gasoline (Borenstein, 1991) and obstetric services (Amin et al., 2004). Due to the rise of the Internet and e-commerce in the past decades, firms are able to price discriminate today more than ever. Consumers can get differential prices based on their past purchasing behavior, web browsing history, demographic characteristics and other information that can be retrieved online. Personalized pricing, which can be seen as an extreme form of third-degree price discrimination, can substantially boost firm's profit (Dubé and Misra, 2017, Shiller, 2020).

Despite the fact that price discrimination theoretically allows a firm to extract additional surplus from the consumers, in practice it can backfire. Perhaps the most famous example is the price testing done by Amazon in 2000, during which the same DVDs were offered at different prices to different users, with the discounts ranging from 20% to 40%. After multiple complaints from the customers who noticed price discrimination, Amazon had to issue an apology and refund thousands of the disadvantaged customers[7], also suffering a damage to its reputation that is difficult to quantify. Another scandal emerged around Adobe, Apple and Microsoft in 2013 as they offered the same digital products to their Australian customers at much higher prices compared to the US market. The three tech giants were called to explain their pricing policy in front of a parliamentary committee, leading Adobe to eventually cut their prices in Australia[8]. On the one hand, third-degree price discrimination can apparently seem

[7]T. R. Weiss, "Amazon apologizes for price-testing program that angered customers" (2000). Accessed at `https://www.computerworld.com/article/2588337/amazon-apologizes-for-price-testing-program-that-angered-customers.html` on 09-07-2020.

[8]L. Whitney, "Adobe cuts prices in Australia following price-gouging probe" (2013). Accessed at `https://www.cnet.com/news/adobe-cuts-prices-in-australia-following-price-gouging-probe/` on 09-07-2020.

unfair; on the other hand, it is widely accepted in other contexts. For instance, various discounts for students, elderly and people with lower income are typically perceived as fair (Wu et al., 2012).

Having said that, when is differential pricing still profitable for the firm? This paper investigates optimal third-degree price discrimination on the markets where consumers have preferences for fairness. I developed an application of the behavioral model by Falk and Fischbacher (2006) to a monopolistic market with two types of consumers with different demand elasticities. Consumer preferences involve reciprocity: thereby, a consumer who is charged a higher price might be willing to punish the firm by lowering his demand; and vice versa, reward the firm for being charged a lower price. I assume that consumers evaluate the fairness of a price by comparing it to the price charged to the other consumer on the market, while also considering the other consumer's payoff – the latter allowed me to model the acceptance of price discounts for consumers with lower income.

The main finding of this model is that the fairness preferences of the consumers put a constraint on price discrimination. The equilibrium price charged to the consumer with a lower demand elasticity (the "richer" consumer) is decreasing in the strength of his preference for reciprocity. Since this price is higher than that charged to the "poorer" consumer, the "richer" consumer might negatively react to it, making price discrimination less attractive. The stronger he reacts, the closer the two prices get, with uniform pricing being optimal at some point. This restraint is mitigated if the payoff of the other consumer is also taken into account, e.g. the "richer" consumer can consider the lower payoff of the "poorer" consumer as a justification for the price disparity. Finally, price discrimination is more attractive if the difference between consumers is large, while potential consumer antagonism to price discrimination makes it suboptimal to uniform pricing when the two consumers are similar.

The most important contribution of this study is the investigation of optimal pricing on the markets where consumers have fairness preferences. Relatively few theoretical models have incorporated fairness concerns into the profit maximization problem of the firm. For instance, Destan and Yılmaz (2020) derived optimal non-linear pricing (second-degree price discrimination) under inequity aversion; Selove (2019) studied the impact of fairness concerns on dynamic pricing in a setting where travel is costly for the consumers. Li and Jain (2016) showed that fairness concerns can soften the competition between firms that practice behavior-based pricing. The models closest to my approach were developed by Okada (2014) and Rotemberg (2011). Okada (2014) assumed that consumers care about the difference in their prices and found that a monopolist prefers uniform pricing over price discrimination if the fairness concerns are strong enough. However, without additional assumptions this model cannot explain why consumers would perceive price discounts for people with lower income as fair. I was able to deliver this prediction by incorporating the payoff of the peer into consumer's utility function. The model by Rotemberg (2011) also predicts that firms practice third-degree price discrimination based on income differences, but avoid discrimination based on differences in the elasticity of demand. The key assumption of this model is that a fraction of the firms are benevolent, i.e. some firms have inherent social preferences towards consumers. Thus, the avoidance of price discrimination can be attributed to non-benevolent firms trying to signal their type as benevolent. I abandon this assumption in my model and show that price discrimination can be suboptimal for a firm even from a purely selfish-rational perspective.

This study also contributes to the debate about the relevant reference price. Several studies have attempted to answer what constitutes this reference price. In some cases, consumers may compare a price to the prices they get for competitors' products (Bolton et al., 2003). Another important reference point is the prices faced by the same consumer in the past, or the so-called self/self comparison (Bolton et al., 2003, Herz and Taubinsky, 2018, Xia et al., 2004). Finally, the prices charged to other consumers can be considered as reference prices (self/other comaprison), leading the reader to the empirical literature on the perceived fairness of third-degree price discrimination, which is reviewed in detail in Chapter 3. I chose to model the reference price as the price charged to the other consumer on the same market, following Xia et al. (2004) and Ho and Su (2009) who claim that comparison between peers is stronger that self/self comparison or distributional fairness between a consumer and a firm.

The rest of the chapter is organized as follows. Section 2.2 introduces the model, its equilibrium and comparative statics, and section 2.3 concludes. The proofs of all propositions in this chapter as well as the Mathematica program for calculation of equilibrium prices can be found in Appendix B.

2.2 The model of a monopolistic market with reciprocal consumers

2.2.1 Model specifications

In this section I present a model of third-degree price discrimination among consumers with reciprocal preferences. Consider a monopolist who sells a good to two consumers with linear demands satisfying Spence-Mirrlees conditions. The payoff function of the first consumer is normalized to $u_1(q_1, p_1) = q_1 - \frac{1}{2}q_1^2 - p_1q_1$, and the payoff function of the second consumer is $u_2(q_2, p_2) = a_2q_2 - \frac{1}{2}a_2q_2^2 - p_2q_2$, with $a_2 \in (0, 1]$. The corresponding linear demands are $\bar{q}_1(p_1) = 1 - p_1$ and $\bar{q}_2(p_2) = 1 - \frac{p_2}{a_2}$. This setting could also be interpreted as two markets with a unit mass of consumers with unit demands, where the valuations for the good are uniformly distributed on $[0, 1]$ on the first market and on $[0, a_2]$ on the second market. For simplicity I assume that there are no production costs, so that firm's profit is the sum of the revenues from the two markets: $\pi = p_1q_1 + p_2q_2$.

Throughout the section, I will consider two pricing options for the firm: either to charge the same price for both consumers (uniform pricing) or to set two different prices that maximize firm's profit (third-degree price discrimination).

Under the standard approach, consumers would care only about their payoff. If the firm charges the same price for both consumers, the optimal uniform price will amount to $p^{UNI} = \frac{a_2}{1+a_2}$, and the corresponding profit is $\pi^{UNI} = \frac{a_2}{1+a_2}$. If the firm engages in price discrimination, the optimal prices are $p_1^{PD} = \frac{1}{2}$ and $p_2^{PD} = \frac{a_2}{2}$, and the corresponding profit is $\pi^{PD} = \frac{1+a_2}{4}$. In the standard setting, uniform pricing yields less profit, therefore the firm always prefers price discrimination.

My version of consumer preferences is an application of the reciprocity model by Falk and Fischbacher (2006). I assume that consumer i has the following utility function:

$$U_i(q_i, p_i, p_j) = u_i(q_i, p_i) + \rho_i\Delta_i(p_i, p_j, \gamma)\sigma_i(q_i, p_i, p_j), \tag{1}$$

where

- $i, j \in \{1, 2\}$; p_i and p_j are the prices the firm charges to consumers i and j respectively.

- $u_i(q_i, p_i)$ is consumer i's quadratic payoff function.

- $\rho_i \geq 0$ is the parameter capturing the strength of consumer i's reciprocity preferences.

- $\Delta_i(p_i, p_j, \gamma)$ is the term representing the kindness of the firm towards the consumer, which amounts to the difference between consumer i's maximal possible payoff given his price p_i and a reference payoff:

$$\Delta_i(p_i, p_j, \gamma) = u_i(\bar{q}_i(p_i), p_i) - \left[\gamma u_i(\bar{q}_i(p_j), p_j) + (1-\gamma) u_j(\bar{q}_j(p_j), p_j)\right].$$

  $\bar{q}_i$ and $\bar{q}_j$ correspond to the linear demands predicted by the standard theory. The reference payoff in square brackets is a weighted average of consumer i's maximal possible payoff given the other consumer's price p_j and the other consumer's maximal possible payoff given p_j. The weight $\gamma \in [0, 1]$ reflects how much consumer i cares about the other consumer's payoff. If $\gamma = 1$, the consumer would care only about his own potential payoff if charged the other price. For instance, with $p_i > p_j$ and $\gamma = 1$, the kindness term would constitute a negative difference between consumer i's maximal payoffs under both prices, meaning that consumer i envies consumer j. If $\gamma < 1$, the other consumer's payoff also matters for the decision making. For computational feasibility, I assume the same γ on both markets.

- $\sigma_i(q_i, p_i, p_j)$ is the reciprocity term, or the difference in firm's actual profit and firm's maximal possible profit on market i:

$$\sigma_i(q_i, p_i, p_j) = p_i q_i - p_i \bar{q}_i(p_i).$$

  Thus, depending on the sign of the kindness term, the consumer may be willing to punish or to reward the firm.

This modification of the original model has a specific structure: while a consumer considers the other consumer (his price and payoff) as his reference point, he reciprocates towards another agent – the firm. This structure was motivated by the observed examples of negative reaction to price discrimination. Consider a buyer on Amazon who notices that his price for a good is higher than some other buyer's price, becomes angry and decides not to buy this good. This is unlikely that at the moment of this decision the buyer compares his well-being to the profits earned by Amazon. He probably pays attention to the prices that other buyers get, and his punishment goes towards the firm that set these prices. A horizontal comparison between consumers' prices and payoffs seems more intuitive than a vertical comparison of consumer's and firm's payoffs. In the rest of the section, I present the important properties of this model.

2.2.2 Equilibrium and comparative statics

Proposition 2.1. *Consumers with reciprocal preferences defined by (1) have the following demand functions:*

$$\begin{aligned}
q_1(p_1,p_2) &= 1 - \rho_1 p_1^2 + \frac{\rho_1}{2} p_1^3 + \\
&+ \left(\frac{-2a_2 + a_2\rho_1 - a_2^2\rho_1 - a_2\gamma\rho_1 + a_2^2\gamma\rho_1}{2a_2} + \rho_1 p_2 + \frac{-\rho_1 + \gamma\rho_1 - a_2\gamma\rho_1}{2a_2} p_2^2 \right) p_1; \\
q_2(p_1,p_2) &= 1 - \frac{\rho_2}{a_2} p_2^2 + \frac{\rho_2}{2a_2^2} p_2^3 + \\
&+ \left(\frac{-2a_2 - a_2\rho_2 + a_2^2\rho_2 + a_2\gamma\rho_2 - a_2^2\gamma\rho_2}{2a_2^2} + \frac{\rho_2}{a_2} p_1 + \frac{-a_2\rho_2 - \gamma\rho_2 + a_2\gamma\rho_2}{2a_2^2} p_1^2 \right) p_2.
\end{aligned}$$

These functions are third-degree polynomials in prices. In the firm's optimization problem, profit becomes a fourth-degree polynomial in prices, making the model intractable. I solved the firm's optimization problem numerically with the help of Wolfram Mathematica 9 software. Parameters a_2, ρ_1, ρ_2 and γ ranged on a grid $(0,1] \times [0,5] \times [0,5] \times [0,1]$ with over a million datapoints. The Mathematica code can be found in Appendix B.2; the proofs of all propositions in this section are provided in Appendix B.1.

Proposition 2.2. *For over a million parameter combinations such that $a_2 \in (0,1]$, $\rho_1 \in [0,5]$, $\rho_2 \in [0,5]$ and $\gamma \in [0,1]$, there exists a unique price pair (p_1,p_2) that maximizes the firm's profit.*

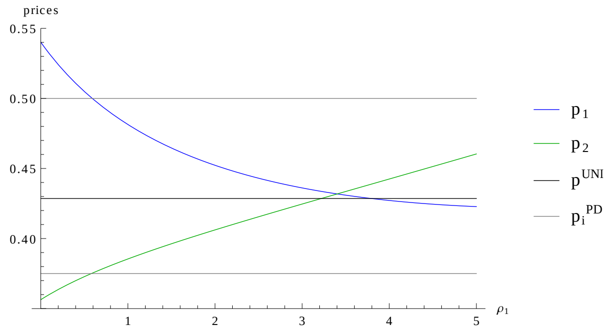
The firm's maximization problem has a unique solution for all datapoints on the chosen grid. I interpolated the resulting prices as functions of parameters. All further propositions in this section are based on the properties of these numerically estimated functions $p_1(a_2,\rho_1,\rho_2,\gamma)$ and $p_1(a_2,\rho_1,\rho_2,\gamma)$. They allow me to examine comparative statics for this model.

Proposition 2.3. *For the following range of parameters, equilibrium prices are weakly decreasing in the reciprocity parameters on the same market and at least weakly increasing in the reciprocity parameters on the other market:*

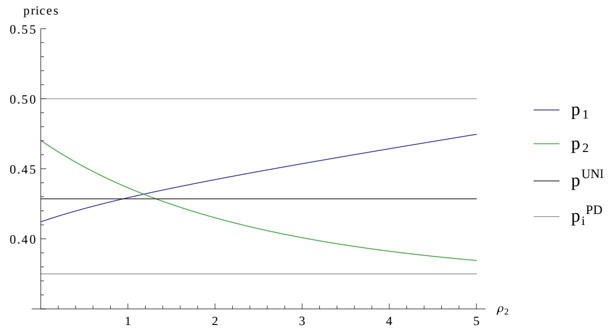
$$\begin{aligned} \textit{if } \gamma > 0 \textit{ and } \rho_i - \rho_j \leq 3.5, \quad & \frac{\partial p_i}{\partial \rho_i} \leq 0; \\ \textit{if } a_2 \geq 0.55, \quad & \frac{\partial p_1}{\partial \rho_2} \geq 0; \\ \textit{for all parameters on the chosen grid}, \quad & \frac{\partial p_2}{\partial \rho_1} > 0. \end{aligned}$$

As an illustration for this proposition, let us consider an equilibrium where the monopolist charges the first consumer a higher price. The first condition on the parameters states that (a) the consumer is not completely altruistic when forming his reference payoff, i.e. considers at least some fraction of his own payoff that he could get given the lower price; and (b) his preference to reciprocate is not too much stronger than that preference of the other consumer. If these assumptions hold, an increase in the reciprocity parameter ρ_1 would imply that the consumer will lower his demand in response to a higher price. Thus, the firm would decrease the price for the first consumer and increase the price for the second consumer, shrinking the magnitude of price discrimination in order to mitigate the negative reciprocal response of the disadvantaged consumer 1.

While it is relatively straightforward why γ should be positive for this to hold, I need to elaborate on assumption (b). If the reciprocity parameter ρ_1 becomes too large, the negative response of consumer 1 becomes so dramatic that it is not profitable for the firm to charge him a higher price. An example in Figure 2.1a shows that at some point firm sets a uniform price on the two markets. Moreover, if ρ_1 exceeds this threshold, the optimal price for the first consumer will be lower than the price for the second consumer. In this case, the first consumer will exercise positive reciprocity by rewarding the firm with an additional increase in demand. As ρ_1 grows, the price p_1 continues to decrease until some point when the firm's benefit from the first consumer's positive reciprocity cannot compensate for the foregone benefit from charging a higher price on a larger market anymore. At this point, the price would actually start to increase in the reciprocity parameter. This situation seems unlikely to happen on a real



(a) Prices as functions of reciprocity parameter ρ_1, with $a_2 = 0.75,\ \rho_2 = 1.5,\ \gamma = 0.8$.



(b) Prices as functions of reciprocity parameter ρ_2, with $a_2 = 0.75,\ \rho_1 = 3,\ \gamma = 0.8$.

Figure 2.1: Equilibrium prices depending on the reciprocity parameters

market, so a boundary on ρ_1 does not sound implausible.

The same logic applies to the comparative statics with respect to the reciprocity parameter on the smaller market. Let us return to an example where the monopolist would charge a lower price for consumer 2. With an increase in ρ_2, the firm could expect a reward from the increased demand of the second consumer. If this consumer has stronger reciprocal preferences, the firm has an incentive to exacerbate price discrimination by lowering p_2 and increasing p_1. The latter only holds if $a_2 \geq 0.55$: if the second market is too small, the benefits from positive reciprocity of consumer 2 do not outweigh the benefits of harvesting the larger market. An illustrative example is given in Figure 2.1b.

As I already mentioned above, as paradoxical as it sounds, it is theoretically possible that it is optimal for the firm to charge a higher price on the smaller market. If the first consumer has very strong reciprocal preferences, his negative response to a higher price would destroy the additional profit of charging him that price. Another interesting observation is that the magnitude of price discrimination can even extend that predicted by standard theory (see Figure 2.1a). This can happen when consumer

1 has weak reciprocal preferences and consumer 2 has strong reciprocal preferences. Thus, the incentive to inflate the price differential because of second consumer's positive reciprocity would outweigh the incentive to shrink it because of first consumer's negative reciprocity. Both situations are probably rare on the real markets, but it is worth for any firm with sensitive customers to think about.

Proposition 2.4. *For a range of parameters, the equilibrium price on the larger market is weakly increasing in the size of the second market. The price on the smaller market is always strictly increasing in its size:*

$$\text{if } \gamma \geq 0.4 \text{ and } \rho_1 \geq \rho_2 \text{ and } a_2 \geq 0.1, \quad \frac{\partial p_1}{\partial a_2} \geq 0;$$
$$\text{for all parameters on the chosen grid}, \quad \frac{\partial p_2}{\partial a_2} > 0.$$

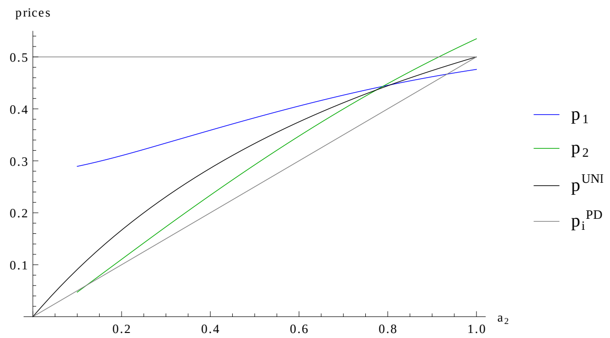


Figure 2.2: Equilibrium prices depending on the size of the smaller market a_2, with $\rho_1 = 3,\ \rho_2 = 1.5,\ \gamma = 0.8$.

Just as with the standard approach, the price on the second market always increases when the market grows. More interestingly, now the optimal price on the first market also depends on the size of the second market. Intuitively, since a larger a_2 always comes with a larger p_2, it creates a higher reference price for the first consumer, making it possible for the firm to increase his price as well without any loss of profit due to negative reciprocity. An example is depicted in Figure 2.2. For the latter to happen, consumer 1 must be substantially selfish in forming his reference payoff, and his reciprocity preference should be stronger than that of the other consumer. Otherwise, the incentive to increase the difference in prices to reap the benefits of the second consumer's positive reciprocity could override the previous argument. Note that this is is only important for relatively small second markets; if the size of the second market is at least half of the first market, p_1 increases in a_2 regardless of the reciprocity parameters. As depicted in Figure 2.2, the difference in prices shrinks as the markets get closer.

Proposition 2.5. *For the following parameters, the equilibrium price on the larger market is weakly decreasing in γ, while the price on the smaller market is weakly increasing in γ:*

$$\text{if } a_2 < 1 \text{ and } \rho_1 > 0, \quad \frac{\partial p_1}{\partial \gamma} \leq 0;$$
$$\text{if } 0.55 \leq a_2 < 1 \text{ and } \gamma \geq 0.2 \text{ and } \rho_2 > 0 \text{ and } \rho_1 - \rho_2 \leq 3, \quad \frac{\partial p_2}{\partial \gamma} \geq 0.$$

To understand the intuition behind Proposition 2.5, let us recall the example with a higher price for consumer 1. If γ increases, his own potentially higher payoff becomes more salient for the reference point than the other consumer's lower payoff. All else being equal, consumer 1 will retaliate more by lowering his demand, and the firm will decrease his price in response. A sufficient condition for this to happen is the existence of reciprocal preferences on the side of consumer 1 (positive ρ_1) and at least some difference in size of the two markets.

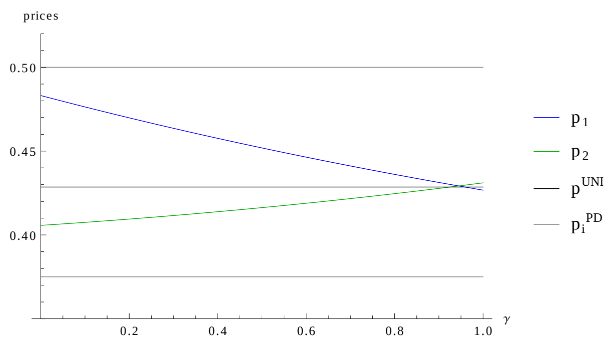


Figure 2.3: Equilibrium prices depending on the selfishness parameter γ, with $a_2 = 0.75$, $\rho_1 = 3$, $\rho_2 = 1.5$.

Similar logic applies to the price on the second market: increasing γ means that the second consumer pays more attention to the fact that his price is lower and is more willing to reward the firm for this, creating room for an increase in his price. The assumptions are stronger though: the second market should be large enough; consumers should not be too altruistic when thinking about their reference points; consumer 2 should have reciprocal preferences; and, finally, the reciprocity parameter of the first consumer should not be too high (an upper bound slightly more restrictive than in Proposition 2.3).The intuition behind these assumptions has largely been explained above. In addition to this, if γ happens to be low and the reciprocity preferences on both sides strong enough, the incentive to shrink the price differential might fade. That is, for instance, an altruistic first consumer could see a higher price as an advantage because the second consumer has much lower payoff anyway. In this case he could be willing to reward the firm instead of punishing it, making the effect of γ on his price obscure. Figure 2.3 provides an example illustrating Proposition 2.5.

2.3 Conclusion

In this study, I developed a model of a monopolistic market with two consumers with reciprocal preferences and derived the optimal third-degree price discrimination strategy. I showed that fairness concerns can have a negative impact on the profitability of price discrimination, with uniform pricing being optimal if the consumers' attitude to (negative) reciprocity is strong enough. If the consumers are aware of each other's payoff, this might mitigate their reaction to price discrimination and give the monopolist more room for the extraction of consumer surplus. On top of that, the attractiveness of third-degree price discrimination shrinks if the markets are of similar size: in this case, the costs imposed by negative reactions of the consumer who gets a higher price outweigh the benefits associated with differential pricing.

This model is a parsimonious setup aimed only at modeling consumer antagonism to third-degree price discrimination in a one-shot market interaction. It remains an open question how optimal price discrimination evolves in a dynamic environment with fairness-concerned consumers. Another interesting direction for future research is modeling competition on such markets: in that case, an angry customer could punish the firm not only by lowering his demand, but also by switching to the competitor's product. Last but not least, apart from third-degree price discrimination, there are other pricing policies that allow to extract consumer surplus, but are still largely perceived as fair: consider, for instance auctions and non-linear pricing. It would be interesting to learn how fairness preferences of the consumers shape the firm's choice between these different pricing options.

The simplicity of this model makes it easy to test its predictions. The next chapter introduces a laboratory experiment designed to test this model and provides the estimates of the model's parameters.

3 Optimal third-degree price discrimination and reciprocity: Experimental evidence

3.1 Introduction

When is a price fair? A vast strand of empirical literature in marketing research was dedicated to the determinants of price fairness or unfairness. Kahneman et al. (1986b) were the first to establish the "dual entitlement" principle: while consumers are entitled to their reference price, firms also have an entitlement to their reference profit. The latter implies that it is fair for firms to raise prices following an increase in costs (see also Bolton et al., 2003, Gielissen et al., 2008, Urbany et al., 1989); however, a firm is not obliged to lower its price if it faces a decrease in costs (Kahneman et al., 1986a). Otherwise, if there are no cost fluctuations, the perception of price fairness crucially depends on the motive of the firm. For instance, consumers see it as unfair if the firm increases prices to take advantage of an increase in demand (Campbell, 1999, Frey and Pommerehne, 1993). On the other hand, higher prices might be deemed fair if the firm engages in charity (Campbell, 1999) or corporate social responsibility (Habel et al., 2016). Lower prices for students, elderly and poorer agents are also widely accepted (Gielissen et al., 2008, Wu et al., 2012).

One of the pricing policies that provokes fairness concerns especially strongly is third-degree price discrimination. Due to the social comparison among the consumers, third-degree price discrimination in its many forms can be perceived as unfair and even become detrimental to firm's profit. A number of survey studies from marketing research detected this phenomenon in various contexts. Huppertz et al. (1978) examined adverse effects of geographic price discrimination: if an item was sold at higher price in an on-campus store compared to off-campus store prices, students were willing to leave the store. Darke and Dahl (2003) discovered antagonism to price discrimination in the form of price discounts. A specific type of discounts, targeting pricing, or offering lower prices to new or regular customers, has also been observed to affect fairness judgement and purchase intentions of the customers (Feinberg et al., 2002, Tsai and Hsiao-Ching, 2007, Wang and Krishna, 2012). Various forms of third-degree price discrimination on the Internet, such as dynamic posted pricing, online couponing, random discounting, price discrimination based on price sensitivity and other Internet-enabled buyer identification techniques have also provoked consumer antagonism, questioning the profitability of these practices (Garbarino and Maxwell, 2010, Grewal et al., 2004, Haws and Bearden, 2006, Huang et al., 2005, Shor and Oliver, 2006).

Although survey studies deliver valuable insights into the perceptions of fairness of price discrimination, the generalizability of their results is limited. We tested consumer antagonism to third-degree price discrimination in a laboratory experiment and estimated the model of price discrimination introduced in Chapter 2. The subjects were matched into markets of three (two consumers and a firm), and made their respective choices: firms could choose between price discrimination or a uniform price, and consumers decided how much to buy from the firm. We had three treatments in our experiment. In the control group (*No Info* treatment), consumers learned neither the price nor the payoff of the other consumer on the market, thus making it optimal for the firm to choose profit-maximizing price discrimination. The price charged to the other consumer was made observable in *Part Info* treatment, designed to trigger reciprocal responses to differential prices. Finally, in addition to the price, in *Full Info* treatment consumers were aware of the other consumer's payoff. The model in Chapter 2 predicted negative reciprocal reactions of lower magnitude in this treatment, imitating the phenomenon of student/elderly discounts. We estimated the parameters of the model in all three treatments and discussed the optimal pricing choices of the firms.

Contrary to previous experimental studies (Englmaier et al., 2012, Leibbrandt, 2020), we find no effect of information treatments on consumer behavior. The subjects in the role of consumers showed no reciprocal reactions to price discrimination and were on average selfish-rational, with only minor deviations from this benchmark. This somewhat surprising result might be attributed to our experimental design: we used the strategy method and repeated games in order to get enough data points to structurally estimate the model. Such an environment might have led our subjects to make more rational decisions instead of acting emotionally, while it is known that emotions play an important role in perceived price fairness (Bougie et al., 2003, Campbell, 2007, Wu et al., 2012). The estimated reciprocity parameters of the model were close to zero, and all deviations from selfish-rational demand were absorbed by the error term. Interestingly, the firms anticipated consumer behavior fairly well. The share of price discriminating firms ranged between 63 and 71% in different treatments and was not significantly different from the best response to actual consumer demands.

Our research adds to the few experimental studies of consumer reactions to third-degree price discrimination. The earlier laboratory experiment by Englmaier et al. (2012) showed that consumers buy less that predicted by standard theory if they are charged a higher price compared to another consumer. The same effect was observed by Leibbrandt (2020) in a lab experiment, while at the same time consumers increased their demand in response to lower discriminatory prices. Allender et al. (2020) also established disadvantaged peer-induced fairness concerns in response to personalized prices in the lab. Finally, Anderson and Simester (2008) investigated consumer reactions to higher prices for larger sizes of women's clothing in a field experiment. They found that the consumers decreased their demand in response to the price premium, leading to a decrease in gross profit of the firm.

We also add to the empirical literature searching for optimal price discrimination schemes. Leslie (2004) was one of the first to quantify the benefits of third-degree price discrimination in a field experiment, estimating the profit gain from discrimination to be about 5% relative to uniform pricing. More recent studies estimated the additional profit associated with personalized pricing on the Internet to be between 10 and 13% (Dubé and Misra, 2017, Shiller, 2020). The consumers in these experiments were presumably unaware of the price differences, otherwise the estimates could have been different. In fact, a laboratory experiment by Allender et al. (2020) revealed that sellers are willing to obfuscate price discrimination to avoid the negative consumer reactions. Several field studies have focused on consumer antagonism to price variation and report mixed results (Anderson and Simester, 2010, Courty and Pagliero, 2010, Sonnabend, 2016). In two laboratory experiments similar to our setting, Englmaier et al. (2012) and Leibbrandt (2020) investigated the attractiveness of third-degree price discrimination for the subjects in the role of the firms. In their earlier experiment, Englmaier et al. (2012) found that the theoretically optimal[9] price discrimination scheme is worse for the firm than a a weaker price difference and even uniform pricing. By design, in this experiment the firms did not have a real choice between the different pricing schemes, and it was not possible to track whether they anticipated optimal prices on their markets. Leibbrandt (2020) reported that in a similar setting sellers tended to avoid price dis-

[9]According to the standard approach without social preferences on the side of the consumers.

crimination; however, its overall attractiveness was partly sustained due to consumers' positive reciprocal responses to lower discriminatory prices. Neither of the above two experiments allowed to quantify the exact benefits of price discrimination. Our study is the first one to estimate the optimal pricing strategy for the firm given the behavioral attitudes of its consumers, and to observe whether the subjects in the role of the firm follow this strategy.

The rest of the paper is organized as follows. Section 3.2 introduces the predictions of the model for our experiment. Experimental design is described in Section 3.3. Section 3.4 presents the empirical results, including treatment comparisons and structural estimation of the model. It is followed by the discussion of the results in Section 3.5, and Section 3.6 concludes. The theoretical foundations for the hypotheses, instructions for the experiment, screenshots of the subjects' computer screens, and the details of our estimation strategy can be found in Appendix C.

3.2 Hypotheses

The model in Chapter 2 provides several predictions for the behavior of consumers and firms in our experiment (algebraical proofs of the hypotheses stated in this section can be found in Appendix C.1). Recall that the experiment consists of three treatments. In *Full Info* treatment, consumers know everything about the other market except for the actual quantity that the other consumer decides to buy. In *Part Info* treatment, the information about the other price is available, but the other payoff function is not revealed. We assume that in this case $\gamma = 1$ since the other payoff is unknown and cannot play a role for decision making. In *No Info* treatment, consumers learn neither the price nor the payoff of the other consumer. We assume that in this treatment the reciprocity parameters ρ_i are equal to zero for both consumers, as if they behaved completely selfish-rationally. We acknowledge that these assumptions on the parameters might be restrictive, however, they are the only plausible assumptions within the framework of the model.

In all treatments in our experiment, the firm cannot set the prices arbitrarily, but has a choice between two options instead: either to set a uniform price p^{UNI} for both consumers or to set two different prices (p_1^{PD}, p_2^{PD}) – a higher price for consumer 1 and a lower price for consumer 2. Each of the two pricing schemes is optimal given the standard selfish-rational preferences and the size of the smaller market a_2. More precisely, the firm chooses between

$$p^{UNI} = \frac{a_2}{1+a_2} \quad \text{and} \quad (p_1^{PD}, p_2^{PD}) = \left(\frac{1}{2}, \frac{a_2}{2}\right).$$

For optimal price and demand functions to satisfy the properties claimed in this section, the market size parameter a_2 was drawn from $[0.5, 1)$ in the experiment. Let us now explore what treatment differences the model predicts.

Hypothesis 3.1. When the firm practices price discrimination, the average demand of consumer 1 in *Part Info* treatment is lower than in both *No Info* and *Full Info* treatment. The average demand of consumer 1 in *Full Info* treatment is lower than in *No Info* treatment if the consumers are selfish enough in forming their reference point:

$$q_1^{PD}(Part\ Info) \leq q_1^{PD}(Full\ Info) \leq q_1^{PD}(No\ Info) \quad \text{if } \gamma \geq \frac{1}{4-a_2}.$$

In the price discrimination scheme available in the experiment, the price for consumer 1 is always higher than the price for consumer 2. It triggers negative reciprocity on the side of the first consumer. In *Part Info* treatment he learns that his price is higher and he could get a better payoff with the other lower price, so he punishes the firm by lowering his demand in comparison to *No Info* treatment rational benchmark. This difference increases in the magnitude of his reciprocal preferences represented by parameter ρ_1; if an average $\rho_1 = 0$, we would observe no treatment effect. In *Full Info* treatment the first consumer additionally learns that the payoff of the second consumer is low. If he cares at least a bit about the other consumer's well-being (reflected in parameter γ), the magnitude of his reciprocal reaction would shrink and thus the demand would be higher in comparison to *Part Info* treatment. If $\gamma = 1$, consumer 1 would not care about the lower payoff of consumer 2, and we would not observe this treatment difference. Finally, let us compare the behavior of the first consumer in *Full Info* and *No Info* treatments. When given complete information, on the one hand, he might be angry about the lower payoff that he gets with his higher price. On the other hand, he might feel satisfied because he is aware that the other consumer gets a lower payoff even when given a lower price. If consumer 1 cares about the other payoff a lot, he could even reciprocate positively towards the firm and increase his demand in comparison to *No Info* treatment. Otherwise, if he pays enough attention to his own payoff ($\gamma \geq \frac{1}{4-a_2}$), the demand in *Full Info* treatment will be lower than in *No Info* treatment. The difference between *Full Info* and *No Info* treatments is again powered by the magnitude of reciprocal preferences; if consumers tend to be selfish-rational, we would observe no treatment effect.

Hypothesis 3.2. When the firm practices price discrimination, the average demand of consumer 2 in *Part Info* treatment is higher than in both *No Info* and *Full Info* treatment. The average demand of consumer 2 in *Full Info* treatment is higher than in *No Info* treatment if the consumers are selfish enough in forming their reference point:

$$q_2^{PD}(No\ Info) \leq q_2^{PD}(Full\ Info) \leq q_2^{PD}(Part\ Info) \quad \text{if } \gamma \geq \frac{a_2}{4a_2 - 1}.$$

This hypothesis represents analogical reasoning behind the behavior of consumer 2, who is triggered to engage in positive reciprocity when given a lower price. As before, theoretically it might happen that his demand in *Full Info* treatment is lower than in *No Info* treatment: if he thinks about the other consumer's higher payoff a lot, he might want to punish the firm even though it is not the firm's fault. However, if consumer 2 pays more attention to his own payoff ($\gamma \geq \frac{a_2}{4a_2-1}$) or, in other words, the price comparison is more salient to him than the payoff comparison, his demand in *Full Info* treatment should be higher than in *No Info* treatment. Treatment effects would also exist only if consumer 2 has any preference for reciprocity, or $\rho_2 > 0$.

Hypothesis 3.3. When the firm chooses a uniform price, there is no difference between the average demands of consumer 1 in *Part Info* and *No Info* treatments. The average demand of consumer 1 in *Full Info* treatment is higher than in *Part Info* and *No Info* treatments:

$$q_1^{UNI}(No\ Info) = q_1^{UNI}(Part\ Info) \leq q_1^{UNI}(Full\ Info).$$

Hypothesis 3.4. When the firm chooses a uniform price, there is no difference between the average demands of consumer 2 in *Part Info* and *No Info* treatments. The average demand of consumer 2 in *Full Info* treatment is lower than in *Part Info* and *No Info* treatments:

$$q_2^{UNI}(Full\ Info) \leq q_2^{UNI}(Part\ Info) = q_2^{UNI}(No\ Info).$$

If the firm decides to set the same price on both markets, it would not provoke reciprocity in the absence of the information about the payoffs. Thus, the demands of both consumers should not differ between *Part Info* and *No Info* treatments. However, the model predicts that alone the information about the other consumer's payoff in *Full Info* treatment might trigger a response even in the absence of difference in prices. Consumer 1 has a higher payoff and therefore might feel advantaged, rewarding the firm with a higher demand; and vice versa for consumer 2. The size of this effect depends on how much consumers care about each other's payoffs. If $\gamma = 1$, there would be no treatment differences whatsoever. However, if any difference is observed, we would not expect it to be large since the firm is not responsible for the difference in consumer payoffs, and consumers would probably anticipate it.

Hypothesis 3.5. The share of the firms that prefer price discrimination over uniform pricing in *Part Info* treatment is lower or equal to that in *Full Info* treatment. The share of price discriminating firms in *No Info* treatment is higher or equal to those in the other two treatments:

$$\frac{N^{PD}}{N}(Part\ Info) \leq \frac{N^{PD}}{N}(Full\ Info) \leq \frac{N^{PD}}{N}(No\ Info).$$

We expect the firms to anticipate the potential reciprocal reactions of the consumers and adjust their pricing strategies accordingly. The model predicts no consumer reaction to price discrimination in *No Info* treatment, thereby all rational firms would prefer it over the uniform pricing, and the expected rates of price discrimination would be highest in this treatment. Due to reciprocal reactions in *Part Info* treatment, price discriminating firms can expect lower demands on the first market and higher demands on the second market. If we assume that consumer 1 has stronger reciprocal preferences ($\rho_1 > \rho_2$), the effect of negative reciprocity will prevail, and thus price discrimination will be strictly less profitable in *Part Info* treatment than in *No Info* treatment. This assumption is warranted since previous studies have found that negative reciprocity is (at least weakly) stronger than positive reciprocity (see, for instance, Al-Ubaydli et al., 2010, Baumeister et al., 2001, Ordóñez et al., 2000).

In *Full Info* treatment, available information about the payoff of the other consumer mitigates their reciprocal reaction to price discrimination. If consumers pay at least some attention to the other payoff ($\gamma < 1$), price discrimination would turn out more profitable for the firm than in *Part Info* treatment, all else being equal.

Finally, the model delivers no clear statement about the difference in the profitability of price discrimination in *Full Info* versus *No Info* treatment. While negative reciprocity on the first market decreases attractiveness of price discrimination, positive reciprocity on the second market and potential altruism on the first market increases its attractiveness. Which of these effects beats the others, is an empirical question. However, note that the frequency of price discrimination in *Full Info* treatment can be at most equal to that in *No Info* treatment since the latter is equal to 1.

3.3 Experimental design

Our laboratory experiment modelled a monopolistic market with a possibility of third-degree price discrimination. In the beginning of the experiment, each subject was randomly assigned one of the three roles: consumer 1, consumer 2 or a firm. This role was retained until the end of the experiment. Subjects interacted on markets of three consisting of a consumer 1, a consumer 2 and a firm.

As the reader already knows, the experiment had three treatment variations. In *No Info* treatment, the subjects in the role of consumers did not have any information about the other consumer on their market. In *Part Info* treatment, prior to making the decision, the consumers learned the price that the firm could set for the other consumer. In *Full Info* treatment, the consumers were aware of both the price and the payoff function of the other consumer on the market. Each subject participated in only one treatment.

Before the beginning of the experiment, subjects received paper-based instructions, which were the same for everyone within each treatment (instructions can be found in Appendix C.2). They were followed by on-screen instructions providing a detailed description of the screen at the moment of making the decision. The on-screen instructions were specific for each role; for the appearance of the screens, refer to Appendix C.3. When all subjects were finished with the on-screen instructions, they started a trial period of market interaction. The structure of this period was identical to the market interaction in the main part of the experiment and served the purpose of familiarizing the subjects with the decision making environment. There was in fact no interaction between the subjects in the trial period since the decisions of the other two market participants were randomly generated for each subject. The payoff earned in this period was not relevant for payment in the experiment.

After the trial period, the main part of the experiment began. It consisted of 20 repeated periods of market interaction. In every period, subjects in the role of consumer 1 and consumer 2 decided how much of a virtual good they wanted to buy from the firm on their current market. The subjects in the role of a firm chose one of the two available price pairs that they could charge the consumers on their market. The decisions were made simultaneously. The detailed procedure is described below.

Decision of the consumers. Each consumer was imposed a payoff function that determined his potential earnings in the experiment. The payoff functions were given by the following formulas:

$$u_1 = \max\left\{100q_1 - 0.5q_1^2 - p_1q_1, 0\right\},$$
$$u_2 = \max\left\{100a_2q_2 - 0.5a_2q_1^2 - p_2q_2, 0\right\},$$

rounded to the next integer, where q_1 and q_2 are the quantities consumer 1 and consumer 2 buy from the firm. The payoffs were presented to the subjects in table form (see Appendix C.3). Consumers could pick the quantities from 0 to 80 with a step of 5. The market size parameter a_2 was randomly drawn from $[0.5, 1)$. We used the strategy method and asked the consumers to make their purchasing decision for both prices that the firm could currently charge them.

On his decision screen, each consumer had two tables corresponding to his payoffs depending on his demand quantities 0, 5, ..., 80 and the two prices the firm could set for him.

- In *No Info* treatment, this was the only information available.
- In *Part Info* treatment, the two prices that the other consumer could get, were also displayed on the screen.
- In *Full Info* treatment, in addition to the prices of the other consumer, his payoff

tables were visible on the screen.

Both consumers were aware of the payoff structure of the firm.

Decision of the firm. Every firm chose a price pair (p_1, p_2) from two available options. These options correspond to the optimal uniform price and the optimal price discrimination scheme, given that consumers care only about their own payoffs:

$$(p_1^{UNI}, p_2^{UNI}) = \left(\frac{a_2}{1+a_2}, \frac{a_2}{1+a_2}\right) \quad \text{and} \quad (p_1^{PD}, p_2^{PD}) = \left(\frac{1}{2}, \frac{a_2}{2}\right),$$

all prices rounded up to the next integer. Depending on the quantities the consumers actually decided to buy for these prices, the payoff of the firm was determined by the sum of the expenses of both consumers:

$$\pi = p_1 q_1 + p_2 q_2.$$

When making the decision, the firm observed four payoff tables on its screen, corresponding to the two consumers on the market and the two pricing schemes the firm could choose from. An assisting calculator for estimating the profit was also available on the screen.

Feedback. When all three market participants made their decisions, the payoffs were calculated and a short feedback was provided. On the feedback screen, consumers saw which price the firm decided to charge them and what their own resulting payoff is (the respective entry in their own payoff table given their chosen quantity). They learned neither how much the other consumer on the market decided to purchase nor how much he earned. The firm learned the consumer quantities for the price scheme it chose, and the resulting profit. It did not learn how much the consumers would have bought if the other price pair was chosen.

After one period had ended, a new period began with the same rules. The roles of the subjects remained unchanged. The payoff function of consumer 1 stayed the same throughout the experiment. The parameter a_2 was randomly drawn in every new period, so the payoff of consumer 2 and the prices changed accordingly. By construction, the payoff of consumer 1 was always higher than the payoff of consumer 2, given the same prices and quantities. Respectively, $p_2^{PD} < p^{UNI} < p_1^{PD}$. The markets were randomly rematched within matching groups of 12 subjects (stranger matching protocol). There were 20 identical periods in the experiment, only one of them being relevant for payoff to avoid income effects. The period to be paid out was randomly chosen in the end of the experiment. The subjects were aware of this procedure as it was explained in the paper-based instructions.

In the end of the final period, we elicited the beliefs about other prices and payoffs on the market in that period. Finally, the experiment was completed with a short questionnaire on sociodemographic characteristics, current major, some school outcomes, mood during the experiment, the experience of participating in economic experiments, etc.

The experiment was conducted in May and June, 2018 at the economic laboratory MELESSA (Munich Experimental Laboratory for Economic and Social Sciences) at the University of Munich, Germany. The experiment was computerized using zTree software (Fischbacher, 2007). We conducted 4 sessions in each treatment, with 24 subjects each. Thus, we had 288 participants in 12 sessions, mostly undergraduate students at the University of Munich. The sessions lasted around 90 minutes, and the average earnings amounted to 20.8 Euros including 6 Euros show-up fee.

3.4 Data analysis

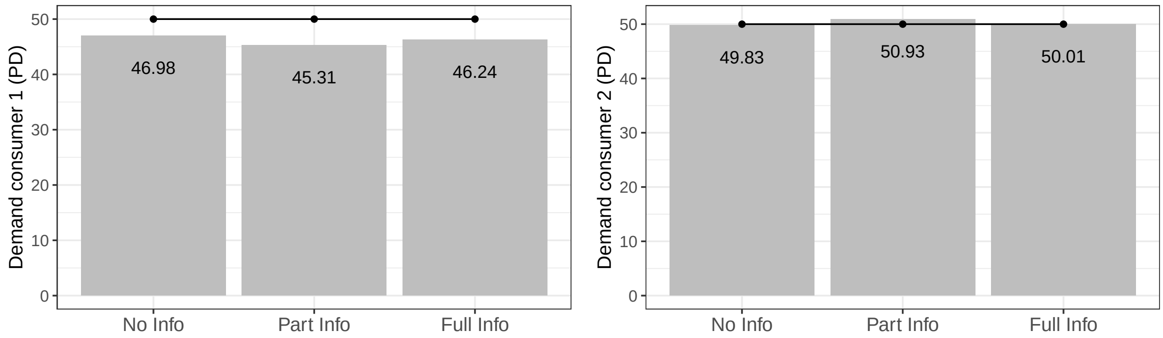
3.4.1 Treatment effects

We start our analysis by comparing the behavior of consumers and firms across different treatments. Our claims are supported by nonparametric statistical tests, which were carried out on the data collapsed on matching group level. Thus, there were 8 independent observations in each treatment and 24 independent observations in total.[10]

Result 3.1. When facing a higher price than the other consumer, the subjects in the role of consumer 1 purchased approximately the same amount of the virtual good in all

[10]Due to a technical error, in one session in *No Info* treatment subjects saw on-screen instructions two times. The instructions were type-specific, and some subjects happened to receive the instructions for two different types of players. These subjects could have inferred something about the other market participants. When we exclude this session from the analysis (2 independent observations), the outcomes of the statistical tests remained unchanged.

three treatments (Kruskal-Wallis test; $p = 0.5778$).



(a) Demand of consumer 1 given p_1^{PD}. (b) Demand of consumer 2 given p_2^{PD}.

Figure 3.1: Consumer demands under price discrimination across the three treatments. The black line corresponds to the demand as predicted by standard theory.

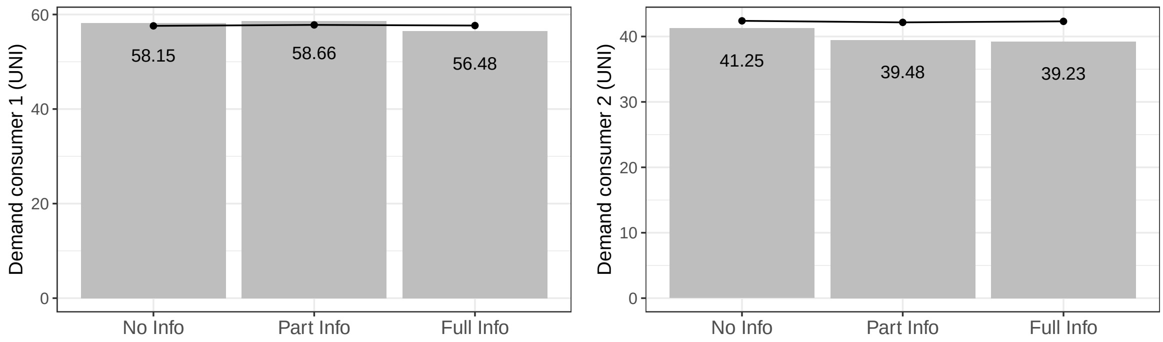
We expected that our treatment variations would trigger antagonism to price discrimination on the side of consumer 1 (see Hypothesis 3.1), but there is no treatment difference to support this hypothesis. We observe neither a reaction to the revelation of the other price nor a reaction to the information about the other consumer's payoff. The average demand of consumer 1 is depicted in Figure 3.1a.

Note that the demand in all three treatments was significantly lower than the standard theoretical prediction in the absence of other-regarding preferences (Wilcoxon signed rank test; $p = 0.0355$, $p = 0.0117$ and $p = 0.0117$ in *No Info*, *Part Info* and *Full Info* treatments, respectively). The disadvantaged consumers lowered their demand equally strong in all three treatments – a somewhat surprising result for *No Info* treatment where the other price on the market was not revealed.

Result 3.2. When facing a lower price than the other consumer, the subjects in the role of consumer 2 purchased approximately the same amount of the virtual good in all three treatments (Kruskal-Wallis test; $p = 0.5257$).

Our expectation that price discrimination would provoke positive reciprocity in consumer 2 was not met either: his demand at a lower price was the same across the treatments (see Figure 3.1b). Moreover, the demand in all three treatments was not different from what the standard approach predicts (Wilcoxon signed rank test; $p = 0.7765$, $p = 0.3264$ and $p = 0.8885$). Consumers with lower payoff thus did not seem to be willing to reward the firm for setting a lower price.

Result 3.3. The demand of consumer 1 was lower in *Full Info* than in *Part Info* treatment when the price was the same for both consumers (Mann-Whitney U Test; $p = 0.0352$). We find no other treatment differences in purchasing behavior of consumer 1 under uniform pricing (Mann-Whitney U Test; *No Info* vs. *Part Info* $p = 0.5984$, *No Info* vs. *Full Info* $p = 0.1275$).



(a) Demand of consumer 1 given p^{UNI}. (b) Demand of consumer 2 given p^{UNI}.

Figure 3.2: Consumer demands under uniform pricing across the three treatments. The black line corresponds to the demand as predicted by standard theory.

Hypothesis 3.3 stated that, if any, the difference between the demand of consumer 1 in *Full Info* treatment and other treatments should be positive since he might perceive his relatively high payoff as an advantage attributable to the firm. However, we observe an opposite effect: the provision of additional information about the second consumer's payoff had a negative effect on the first consumer's demand.

The demand of consumer 1 is higher than the rational benchmark in *Part info* treatment (Wilcoxon signed rank test; $p = 0.0497$), but not in the other treatments (*No Info* $p = 0.3270$, *Full Info* $p = 0.4008$). The average demands are presented in Figure 3.2a.

Result 3.4. The demand of consumer 2 was the same in all three treatments when he got the same price as the other consumer (Kruskal-Wallis test; $p = 0.3090$).

In line with Hypothesis 3.4, provision of additional information does not have an effect on the demand of consumer 2 under uniform pricing.

Interestingly, consumer 2 purchased less than predicted by standard preferences in *Part info* and *Full info* treatments (Wilcoxon signed rank test; $p = 0.0173$, $p = 0.0207$). In the treatments with at least some information about consumer 1, uniform pricing has provoked a significant negative response by consumer 2. In *No Info* treatment, the demand was also lower than rational, but not significantly so ($p = 0.1824$; see Figure 3.2b).

Result 3.5. The share of the firms practicing price discrimination was not significantly different across the three treatments (Kruskal-Wallis test; $p = 0.2856$).

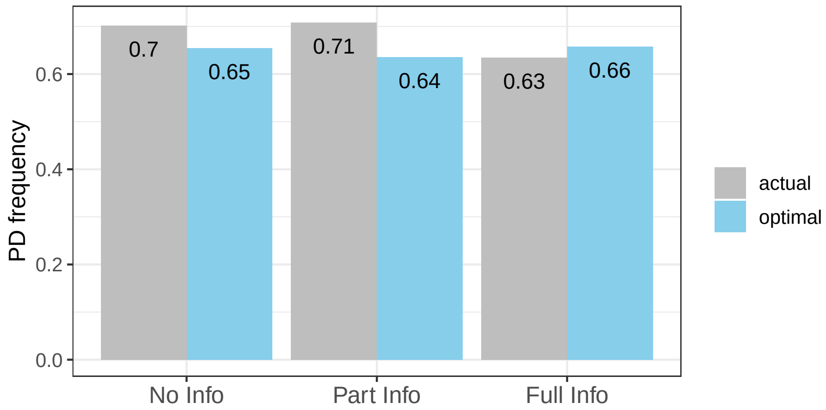


Figure 3.3: Share of price discriminating firms across the three treatments. The gray bars correspond to the actual frequencies of price discrimination, the blue ones – to the expected optimal frequencies given actual consumer demands on each market.

Recall that under standard assumptions about consumer preferences it is always rational for the firm to engage in price discrimination. While we observe no differences in price discrimination rates across treatments, in all treatments its prevalence is significantly lower than 100% (Wilcoxon signed rank test; $p = 0.0117$, $p = 0.0115$ and $p = 0.0116$). Is this behavior optimal given the preferences of the consumers in the experiment? We calculated the optimal response of the firms to the demand of the consumers on every market and plotted it along with the actual frequencies of price discrimination in Figure 3.3. The actual behavior of the firms is not significantly different from the optimal behavior in any of the treatments (Wilcoxon signed rank test; $p = 0.5754$, $p = 0.0929$ and $p = 0.6744$), suggesting that the firms anticipated consumer reactions fairly well.

Finally, we check whether our findings are robust to regression analysis with additional control variables from our experiment. The dependent variables in our models included the demands of both consumers under both pricing schemes as well as the binary firm's choice. As explanatory variables we selected treatment dummies, the market size parameter a_2, current period in the experiment and the following self-reported characteristics from the questionnaire: gender, age, math grade in school, understanding of the experimental rules, major and mood during the experiment. The results are reported in Table 3.1.

First, the regression coefficients are consistent with the outcomes of the comparison tests reported earlier in this section. We observe no treatment differences in the behavior of consumers and firms, apart from the demand of consumer 1 at the uniform price: it is lower in *Full Info* treatment than in both *No Info* and *Part Info* treatments (models 3 and 4 in Table 3.1). This observation is also consistent with Result 3.3, although the nonparametric comparison between *Full Info* and *No Info* did not reach statistical significance probably due to the small amount of independent observations.

Second, let us examine the effect of the market size parameter a_2 on the demands of both consumers. If the consumers were completely selfish and rational, they both would buy 50 units at the discriminatory price regardless of the size of the second market. The demand of consumer 2 at a lower price indeed does not depend on a_2; we also observed in Result 3.2 that his demand was not different from the standard rational benchmark in all three treatments. However, this is not the case for consumer 1 when he faces a higher price: his demand increases in a_2. When a_2 is lower, the difference in prices becomes more pronounced, thus consumer 1 is more likely to punish the firm by lowering his demand. Remarkably, this relationship is observed also in *No Info* treatment, where the consumers were not informed about the other market at all. It is in line with our observation that in all treatments the demand of consumer 1 was lower than the standard prediction (see Result 3.1). Now consider the relationship between the market size a_2 and the demands at the uniform price. The uniform price is increasing in a_2. In the absence of other-regarding preferences, the demand of consumer 1 would therefore decrease in a_2, while the demand of consumer 2 would increase in a_2. We can see in Table 3.1 (models 3-5) that this relationship holds. The effect is quite strong, suggesting that the subjects were fairly rational when making these decisions. The magnitude of the effect is weaker for consumer 1 in *Full Info* treatment – in line with Result 3.3.

In addition, the size of the second market affects the probability of the firm to choose price discrimination. While it should not be the case in the absence of social preferences, in our experiment firms often preferred uniform pricing over price discrimination. Model 6 in Table 3.1 shows that the probability to engage in price

Table 3.1: Consumer demands and firms' pricing

	q1pd (1)	q2pd (2)	q1uni (3)	q1uni (4)	q2uni (5)	firmpd (6)
partinfo	-2.387	1.300	2.578		-3.075	0.737
	(3.542)	(2.053)	(2.538)		(2.431)	(1.095)
fullinfo	-7.240*	2.206	-8.620**	-11.20***	-1.483	0.965
	(4.257)	(2.702)	(3.964)	(3.630)	(2.376)	(1.004)
noinfo				-2.578		
				(2.538)		
a	3.670**	0.000274	-33.26***	-36.59***	33.88***	-3.109***
	(1.492)	(1.823)	(2.180)	(1.817)	(0.698)	(0.939)
partinfo*a	1.792	-0.575	-3.332		2.789	-1.000
	(3.896)	(2.255)	(2.841)		(2.765)	(1.397)
fullinfo*a	8.834*	-2.458	9.059**	12.39***	0.304	-1.596
	(4.518)	(2.894)	(4.237)	(3.997)	(2.802)	(1.266)
noinfo*a				3.332		
				(2.841)		
period	0.0443	0.0706**	0.234***	0.234***	-0.0359	0.0362***
	(0.0416)	(0.0304)	(0.0371)	(0.0371)	(0.0351)	(0.0106)
male	-0.670	2.314***	1.310	1.310	-1.645	0.303
	(0.920)	(0.550)	(0.859)	(0.859)	(1.068)	(0.301)
age	-0.0265	-0.0238	-0.142**	-0.142**	0.0757	-0.0580*
	(0.0752)	(0.0363)	(0.0685)	(0.0685)	(0.0482)	(0.0335)
mathgrade	0.108	-0.128	0.907*	0.907*	-1.081**	-0.0426
	(0.319)	(0.351)	(0.518)	(0.518)	(0.490)	(0.128)
understanding	1.481*	0.660*	-0.725	-0.725	-0.0499	0.489*
	(0.874)	(0.345)	(0.801)	(0.801)	(0.631)	(0.250)
experience	-0.322	-0.549	0.571	0.571	0.355	0.267*
	(0.503)	(0.377)	(0.355)	(0.355)	(0.704)	(0.143)
major_econ	-1.143	0.307	0.335	0.335	-1.069	-0.184
	(1.416)	(0.711)	(0.739)	(0.739)	(1.278)	(0.378)
mood_exp	1.061*	0.464	-0.680	-0.680	0.165	0.131
	(0.618)	(0.422)	(0.459)	(0.459)	(0.766)	(0.142)
Constant	41.31***	48.38***	82.35***	84.93***	16.81***	2.500*
	(3.278)	(2.309)	(3.643)	(3.437)	(2.356)	(1.295)
Observations	1920	1920	1920	1920	1920	1920
R^2	0.0454	0.0712	0.320	0.320	0.297	

Panel regressions with random effects (models 1-5: OLS; model 6: logit).
Cluster-robust standard errors in parentheses (matching group level).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

discrimination decreases in a_2. Given that consumers on the first market were averse to price discrimination to some extent in all treatments, this effect could overweigh the benefits of discrimination when the markets are too close.

Finally, some of the control variables are correlated to the decisions of our subjects. We find learning effects in some of the situations: namely, first consumers increased their demands at the uniform price over time (models 3, 4), second consumers increased their demands at the lower price, and firms discriminated more frequently – all pointing towards the standard rational benchmark. We also considered math grade in school, understanding of the rules of the game, previous experience of participating in economic experiments and being enrolled in either a business or an economics program as proxies for rational behavior. None of them mattered consistently, apart from the understanding of the rules being marginally significant for some choices.

3.4.2 Estimation of the model

We gained important insights into the behavior of our subjects by using non-parametric comparison tests and regression analysis earlier in this section. Non-parametric comparisons lacked power though, and regressions assumed a linear relationship between outcome variables and explanatory variables. Now we are going to estimate our structural model and explore the decisions of consumers and firms in further depth.

To account for any observed behavior not captured by the model, we add an error term to consumer preferences:[11]

$$U_i^{\epsilon}(q_i, p_i, p_j, \epsilon_i) = u_i(q_i, p_i) + \left(\rho_i \Delta_i(p_i, p_j, \gamma_i) + \epsilon_i\right) p_i q_i. \tag{2}$$

The error term allows for deviations from rational behavior that are not explained by the model (for instance, inequality aversion towards the firm). If ϵ_i is negative, consumer i is willing to buy less than the model predicts; if positive, his demand would be higher. We assume that ϵ_i is normally distributed with mean α_i and variance σ_i^2. The assumption about the parameter γ being the same for both consumers is relaxed,

[11]In equation 1 the reciprocation term is defined as the difference between the firm's actual profit and its profit given the consumer's payoff-maximizing demand. Since the latter term does not depend on the actual purchased quantity, it does not matter for the consumer's maximization problem. For simplicity, we drop it hereafter from the utility function.

allowing for different individual levels γ_i.

The estimation procedure finds the error mean α_i, its variance σ_i^2, the reciprocity parameter ρ_i and the "selfishness" weight γ_i that maximize the log likelihood function $\ln L_i(p_i, p_j, \alpha_i, \sigma_i, \rho_i, \gamma_i) = \sum_{k=1}^{40} \ln Pr[q_{ik} = q | p_i, p_j, \alpha_i, \sigma_i, \rho_i, \gamma_i]$ for the observed choices of consumer i in 20 periods for each of the price pairs $(p_i,\ p_j)$. The detailed estimation procedure is described in Appendix C.4. We continue to assume that in *Part Info* treatment subjects were completely selfish in forming their reference point and $\gamma = 1$, leaving us with three parameters to estimate for each consumer. On top of that, in *No Info* treatment we assume that the reciprocity parameters are equal to zero and estimate only the mean and the variance of the error term.

The parameters estimated for each individual consumer are plotted in Figures 3.4, 3.5 and 3.6.

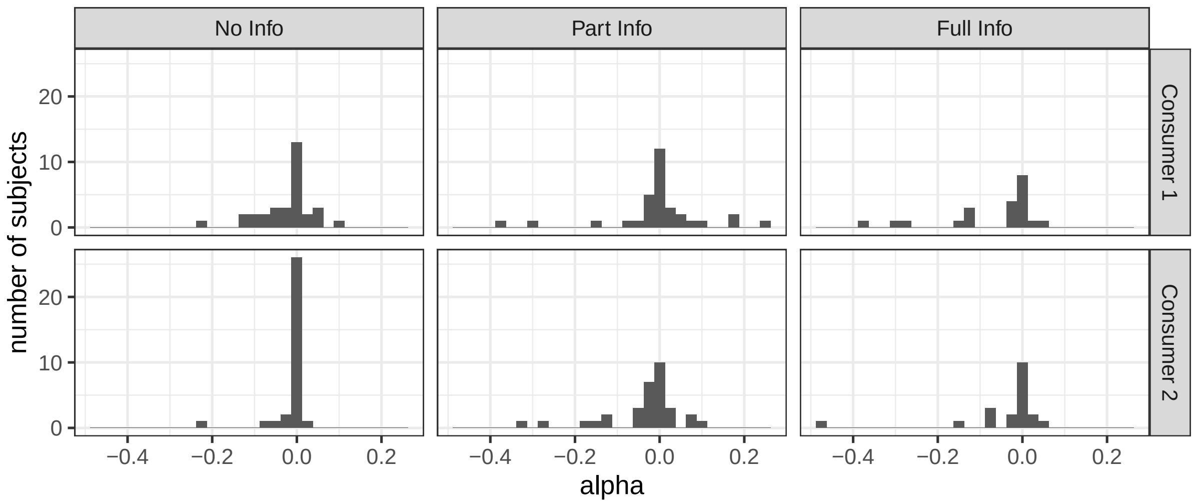


Figure 3.4: The distributions of individual maximum likelihood estimates of the error mean α across consumer types and treatments.

Figures 3.4 and 3.5 show the distributions of the parameters associated with the error term ϵ_i. We observe in Figure 3.4 that in all treatments for all consumer types the error mean α has its mode at zero, but the distributions tend to be negatively skewed. The standard deviation of the error term is presented in Figure 3.5. There is nonnegligible variance of the individual consumer choices, potentially associated with learning. The error is more centered around zero and less dispersed in the *No Info* treatment, meaning that the information treatments added more noise to the choices, which cannot be explained by the model.

Figure 3.6 shows the individual estimates of the fairness parameters. The reciprocity

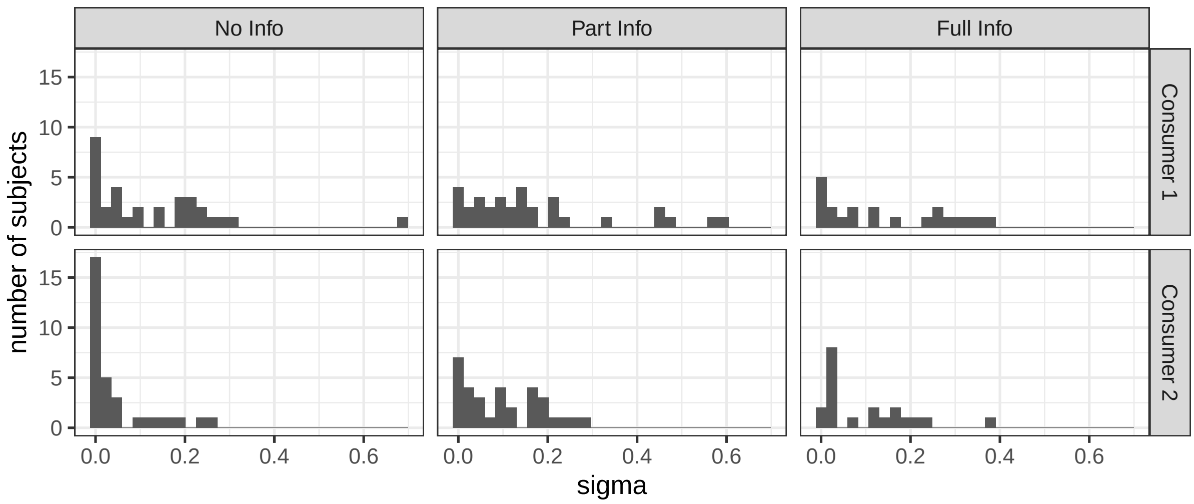
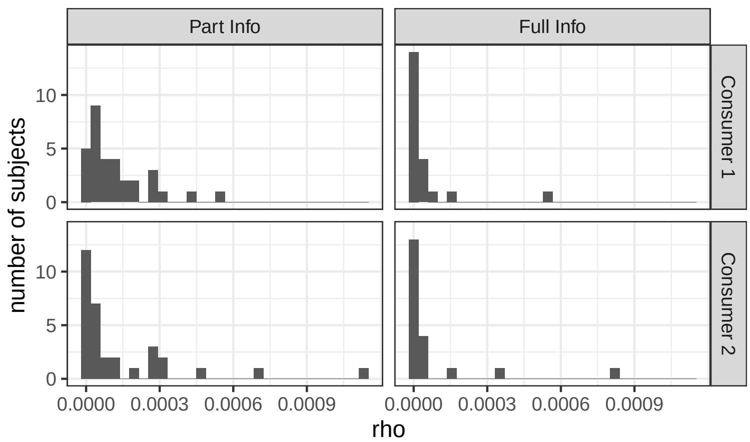
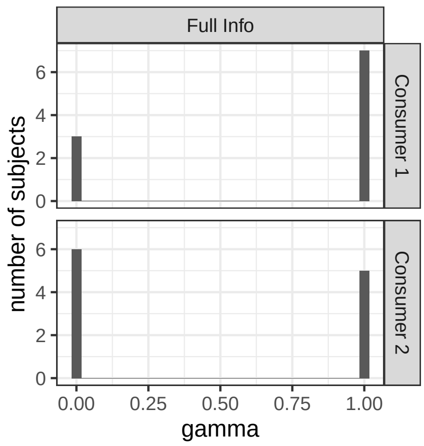


Figure 3.5: The distributions of individual maximum likelihood estimates of the standard deviation of the error σ across consumer types and treatments.



(a) Reciprocity parameter ρ.



(b) "Selfishness" parameter γ.

Figure 3.6: The distributions of individual maximum likelihood estimates of ρ and γ across consumer types and treatments.

parameter ρ is very close to zero for all subjects in the information treatments. It is more dispersed in *Part Info* treatment than in *Full Info* treatment. For those subjects with positive ρ in *Full Info* treatment, we could estimate the weight γ put on consumer's own payoff in his reference point. Two thirds of those subjects in the role of consumer 1 had $\gamma = 1$, thus focusing solely on their own payoff they could get at the lower price (see Figure 3.6b). For the rest of the consumers on the first market γ was approximately equal to zero, putting the whole weight on the lower payoff of the other consumer. The distribution of γ among the subjects in the role of consumer 2 is similar: there are no interior values of the weights; however, there are slightly more subjects concerned with the payoff of the other consumer on the market.

Furthermore, we pooled the data to evaluate the average parameters for each consumer type across different treatments. We had 32 subjects in each treatment in each role, making decisions for two pricing schemes in 20 periods, with a total of 1280 observations per cell. Maximum likelihood estimates are presented in table 3.2.

Table 3.2: Estimated parameters of the structural model

treatment	consumer	α	σ	ρ	γ
No Info	1	-0.0257	0.1770		
		(0.0050)	(0.0036)		
	2	-0.0107	0.0941		
		(0.0027)	(0.0020)		
Part Info	1	0.0044	0.1745	0.0001	
		(0.0063)	(0.0036)	(0.0000)	
	2	-0.0399	0.1223	0.0001	
		(0.0046)	(0.0025)	(0.0000)	
Full Info	1	-0.0581	0.2373	0.0000	0.0000
		(0.0069)	(0.0051)	(0.0000)	(.)
	2	-0.0422	0.1547	0.0001	0.9175
		(0.0097)	(0.0032)	(0.0000)	(0.0967)

The average error mean α has small values, all negative except for the average consumer 1 in *Part Info* treatment. The error was closer to 0 and less dispersed (smaller σ values) for consumer 2 in all treatments. The reciprocity parameter ρ was very close to 0 in all information treatments. Due to its very small value for consumer 1 in *Full Info* treatment, the respective weight γ could not be identified. For consumer 2, γ was close to 1, thus making the subjects primarily concerned with their own potential payoff as their reference point. All in all, the results of the estimation suggest that the consumers were behaving very close to the rational benchmark in our experiment, more so in the role of consumer 2. Any deviations we observe cannot be explained by reciprocity in the framework of the model, but are rather captured by the error term.

3.5 Discussion

We observed that, contrary to our expectations, information treatments did not affect the behavior of consumers in our experiment. They acted very close to the predictions of the standard theory; in some situations consumers did deviate from the rational benchmark and tended to buy less, but it could not be explained by our reciprocity model. Why did we fail to replicate the finding that third-degree price discrimination can trigger reciprocal reactions?

First, our calibration of the model in the experiment often led to large differences between the consumer's and the firm's payoffs. Consumer 1 always earned less than the firm under price discrimination and the majority of uniform prices (in some cases, if the second market was small, he could earn more with a uniform price than the firm). Consumer 2 always earned less than the firm, in some rounds much less if the size of his market was calibrated to be small. The average earnings from the experiment are laid out in Figure 3.7. This inequality was perceived as unfair by some of the subjects: when we asked for informal feedback in the questionnaire at the end of the experiment, roughly 12% of the consumers complained that they were treated unfairly in comparison to the firms. It could have shifted the focus from horizontal comparison between buyers' payoffs to vertical comparison between the seller and the buyer. This speculation is consistent with negative values of the estimated error mean α (see Table 3.2).

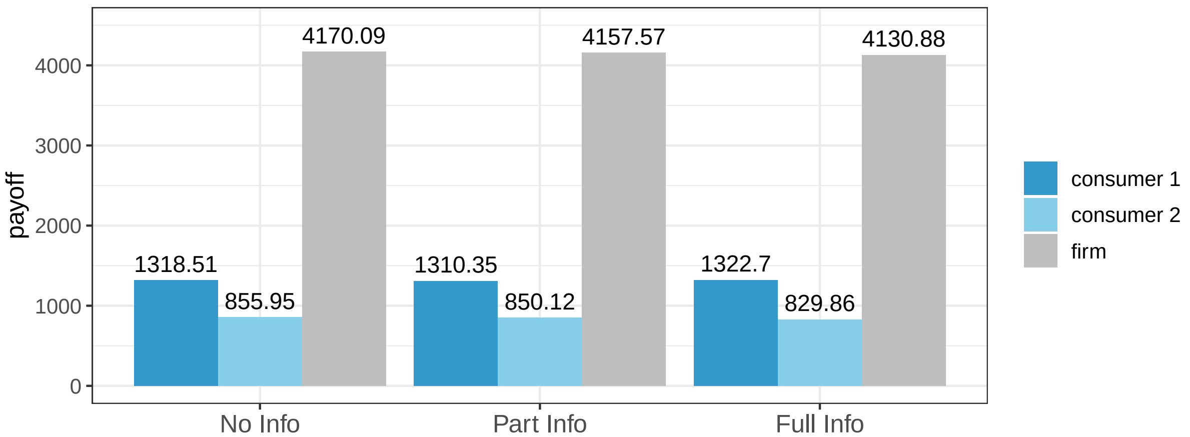


Figure 3.7: Average earnings in the experiment (in experimental currency units)

Thus, we could speculate that inequality averse preferences (for instance, Fehr and Schmidt, 1999) account for our results. However, at this point we should recall the difference in average demands in Figures 3.1 and 3.2. Consumer 1 bought less than predicted by the standard approach in all treatments when facing a higher discriminatory price, but behaved perfectly in line with standard predictions when given the same price. This could still be in line with inequality aversion for very small uniform prices; however, with an average uniform price consumer 1 earned less than the firm. Thus, his average demand at uniform prices should have been lower than selfish-rational, but it was not the case. Moreover, if consumers acted only out of inequality aversion towards the firm, the demand of consumer 2 at each price would be lower than the rational benchmark. This is not the case: consumer 2 bought less for a uniform price, but was selfish-rational with a lower price. This observation contradicts pure inequality aversion since the gap between the payoffs of consumer 2 and the firm was still large under price discrimination, so an inequality averse consumer 2 would be willing to reduce it.

A different choice of the reference point could be another potential explanation for the observed consumer behavior. The model we applied assumed that consumers make self/other comparisons, i.e. compare their own prices to those of the other consumer on the market *within* each pricing scheme. For instance, when consumer 1 decides how much to buy at a higher discriminatory price, the lower price of consumer 2 serves as a reference point for him. Analogously, the reference point for the demand of consumer 1 at the uniform price is this price itself. However, consumers could well compare their own prices *between* pricing schemes (self/self comparisons): i.e. when consumer 1 buys at a discriminatory price, he considers his own potential uniform price as the reference point. In turn, his discriminatory price could be the reference point for his demand at the uniform price.

Although self/other comparisons are considered to have a greater effect on perceived price unfairness in the literature (Xia et al., 2004), self/self comparisons were unintentionally made salient in our experiment. Since we used the strategy method, the consumers made their decisions for two different prices on one screen (see Appendix C.3). Shifting the reference point to one's own price in combination with reciprocal preferences could explain average demands of the consumers (refer back to Figures 3.1 and 3.2): consumer 1 punished the firm for price discrimination since his price was higher than the uniform price, but behaved selfish-rationally under uniform pricing. Respectively, consumer 2 punished the firm for the uniform price since he could get a lower one under price discrimination, and behaved rationally when he got this lower price. The difference between own prices was equally salient in all treatments. It could potentially explain why consumer 1 punished the firm even in *No Info* treatment. Another piece of suggestive evidence can be taken from the regression analysis in Table 3.1: the model in column (1) shows that the demand of consumer 1 at the discriminatory price depends on the size of the second market a_2 in all treatments, including *No Info*. Neither the standard approach nor the model we used predicts that this demand should depend on a_2 in *No Info* treatment. The fact that it does suggests that the uniform price, being the only piece of information about the other market in *No Info* treatment, did affect the demand of consumer 1 at the discriminatory price.

Finally, one feature of our experimental design was the duration and the complexity of the experiment. While the decision was relatively easy for the consumers, the subjects in the role of the firm took more time to figure out the their pricing strategies, this leading to longer waiting times between rounds for the consumers. The decisions were made repeatedly for 20 rounds, the instructions and payoff structures were relatively complex, and many consumers had to wait between rounds – it all made roughly 13.5% of our subjects complain in the questionnaire that the experiment was long, repetitive and boring. Admittedly, it was not the best environment for emotional choices like an angry reaction to a higher price. It likely contributed to the consumer choices being very close to the rational benchmark.

3.6 Conclusion

We tested the predictions of the model in Chapter 2 in a laboratory experiment with three treatments varying the amount of information consumers possess about each other. Contrary to previous empirical studies, we were not able to replicate consumer antagonism to price discrimination. The subjects in the role of consumers displayed very little deviations from selfish-rational behavior, and the respective optimal pricing choices were anticipated by the subjects in the role of the firms.

We chose to employ the strategy method and repeated interactions in our experiment in order to estimate the parameters of the model on every single market, and thus to be able to arrive at the optimal pricing strategy for every firm. Unfortunately, this design choice came at a cost of not finding consumer resistance to price discrimination. It raises the question of whether it is possible for the researchers to estimate optimal third-degree price discrimination at the level of individual firms without distorting consumer reactions to it. First, distributional fairness between the consumers and the firms seemed to play a role in shaping our results, despite our efforts to make it as non-salient as possible. It is not clear to what extent this consideration extrapolates to the real markets, but researchers should be aware of it at least in the laboratory. Second, our experimental design might have induced the consumers to act more rationally. Using direct response instead of the strategy method could provoke more emotional purchasing decisions at the cost of less data at the individual level. Third, along with the strategy method, we in fact practiced dynamic pricing in our experiment by repeating the market interactions, potentially contributing to the relative salience of self/self price comparisons. It delivers an important insight for the firms who intend to estimate the demands of their consumers with the help of dynamic pricing and practice price discrimination at the same time.

Despite the observed rationality of the subjects in our experiment, our approach made another step towards understanding optimal pricing with behavioral constraints. It is left for future research to explore how fairness concerns of the consumers impact the profitability of other pricing practices. For instance, second-degree price discrimination and auctions are also widespread ways to redistribute consumer surplus, while they might be perceived as fair since the consumer has more influence on the price he gets compared to third-degree price discrimination. An empirical investigation of consumer reactions to different pricing practices would contribute to our understanding of the firm's choices between them.

4 The determinants of gender bias in student evaluations of teaching

4.1 Introduction

It has been long known that women fare worse economic outcomes than men on multiple dimensions. The gender pay gap, albeit having decreased over the past years, is estimated to be about 20% in the US (Blau and Kahn, 2017). Between 20 and 30% of all senior management positions globally were held by women in the past decade.[12] Among many other factors contributing to the observed disparities, gender-based discrimination has been offered as an explanation by a broad spectrum of studies. Research has investigated the gender bias against women in employment decisions (Goldin and Rouse, 2000, Neumark et al., 1996), product markets (Ayres and Siegelman, 1995, List, 2004) and politics (Eyméoud and Vertier, 2017). Extensive reviews of experimental literature on gender discrimination can be found elsewhere (Riach and Rich, 2002, Rich, 2014, Bertrand and Duflo, 2017, Neumark, 2018).

One particularly persistent example of the gender gap concerns the position of female faculty in academia. Being a woman is associated with decreased academic rank both in Europe and the United States. In the US men outearn women in all faculty ranks. The current gender pay gap constitutes almost 20% and has not budged over the last ten years.[13,14] Women are less likely to get tenure (Weisshaar, 2017), apply for and receive fewer research grants (Ley and Hamilton, 2008, van der Lee and Ellemers, 2015) and publish fewer papers (Jagsi et al., 2006, Larivière et al., 2013, West et al., 2013). Besides possible differences in the productivity and inherent preferences to work in the competitive environment of the academia, one potential factor contributing to the gender disparities is the gender bias observed in student evaluations of teaching (SET). The evaluations are widely used in hiring and promotion decisions and thus indirectly affect future academic career. Female faculty has consistently been evaluated

[12]Grant Thornton, *Women in Business: Building a Blueprint for Action (2019).* Accessed at `https://www.grantthornton.global/en/insights/women-in-business-2019/women-in-business-report-2019/` on 24-07-2020.

[13]American Association of University Professors, *The Annual Report on the Economic Status of the Profession, 2019–20.* Accessed at `https://www.aaup.org/report/annual-report-economic-status-profession-2019-20` on 27-07-2020.

[14]European Commission, *She Figures 2018.* Accessed at `https://ec.europa.eu/info/publications/she-figures-2018_en` on 27-07-2020.

worse than their male colleagues, observed both in university SET (Wagner et al., 2016, Fan et al., 2019, Funk et al., 2019) and online evaluations (Arceo-Gomez and Campos-Vazquez, 2019, Rosen, 2018). Two recent studies by Boring (2017) and Mengel et al. (2019) found a robust gender penalty in student evaluations of teaching, primarily driven by male students, sustained after controlling for measures of teaching effectiveness such as student grades and self-study hours. This fact combined with women's responsiveness to student feedback (Buurman et al., 2018) might explain why women spend more time on teaching activities and less on research (see Santos and Dang Van Phu, 2019), in turn, impeding their own success in academia.

Is it pure discrimination that drives the observed gender differences in SET, or are there underlying differences in teaching unobserved by the researcher? We attempt to answer this question by proposing a laboratory experiment, which varies the gender of the instructor exogenously and minimizes the differences in teaching style, allowing us to infer whether the discrepancy in evaluations stems from discrimination. The teachers in our experiment will make a video lecture consisting of a short self-introductory video and narrated slides, following the approach by Chisadza et al. (2019). On top of that, we exogenously vary the quality of the lecture to see how evaluations respond to it depending on the gender of the teacher. Additionally, as a way to overcome potential gender bias in SET, we propose a subtle intervention that communicates the effect of evaluations on the teacher's well-being to the students. The teachers will have the possibility to earn a monetary prize based on the student evaluations from the experiment. This possibility will be either not mentioned or explicitly stated to the students before they proceed to the evaluations. Finally, in the controlled laboratory environment we will be able to measure the students' attention to the lecture and the evaluations and observe whether this channel contributes to gender bias in SET.

First, our study will contribute to the experimental literature trying to identify the gender bias attributable to discrimination in SET. While field data is essentially important for policy-making, it is not possible to disentangle discrimination from unobservable differences in teaching quality and/or style in observational data. The famous experiment by MacNell et al. (2015) manipulated the perceived identities of real male and female instructors in an online course. Despite the small sample size, the study detected gender bias in evaluations on several dimensions. However, the experiment was not double-blind, potentially leading the instructors to behave differently when assigned a different displayed gender role. Burnell et al. (2020) found gender bias in evaluations for a hypothetical hiring decision in academia, where the students had to read the text of a potential lecture, only knowing the name of the lecturer. An interaction of gender and age bias was detected in two experiments with narrated slides, which were manipulating the stated identity of the lecturer (Arbuckle and Williams, 2003) or his/her voice (Doubleday and Lee, 2016). Our approach is closest to the deception-free experiment conducted by Chisadza et al. (2019) in South Africa: the authors used narrated slides with a picture of a real lecturer attached to the bottom corner, and manipulated the race and gender of the lecturer. Surprisingly, they did not find gender bias *against* female lecturers; on the contrary, students evaluated women higher than men. The authors speculate that the dominance of female teachers in high school in South Africa might have affected the expectations of the subjects, who were all first-year students in economics. We hope to avoid this effect by recruiting students with diverse educational backgrounds and by selecting the subject of the lecture to be mathematical since women are typically expected to underperform in sciences comparative to men (see Reuben et al., 2014). Another caveat of the study by Chisadza et al. (2019) lies in the treatment variation of the lecturer's gender: since there was only one lecturer per treatment, the treatment differences might have been driven by the differences in the instructors' personal characteristics. In fact, the treatment effects vanish when authors control for teacher's accent. We intend to recruit 4 different lecturers in each treatment condition in order to reduce the variance in their verbal and non-verbal attributes of teaching.

Second, to our knowledge, our study is the first to measure the response of the gender bias in SET to exogenous variation in the quality of teaching. Instead of relying on proxies of teacher's productivity such as student performance, we can directly control for teaching quality and estimate its effect on the evaluations depending on the gender of the lecturer. On the one hand, we will learn whether it is good or bad teachers who suffer from the gender bias in SET. On the other hand, our experiment contributes to the debate about the validity of SET as a measure of teaching effectiveness. While still being widely accepted in academia, student evaluations of teaching are criticized by a growing body of literature. One can think of the famous Dr. Fox effect (Naftulin et al., 1973) as one of the first hints that student evaluations do not necessarily reflect the real quality of teaching. In fact, the Dr. Fox experiment was extended by Ware and Williams (1975) who hired the same actor to perform a lecture under different treatment conditions: one dimension varied the quality of the content, and the other dimension varied the seductiveness of the lecturer. The result was disturbing: content affected the evaluation scores only under low seductiveness, while a highly seductive teaching style received high ratings irrespective of the content. Moreover, experiments have shown that judgements of very short content-free videos of the lecturers (Ambady and Rosenthal, 1993, Babad et al., 2003) or first impressions of their personality after a brief encounter (Clayson and Sheffet, 2006) can predict a major fraction of variance in the subsequent SET. The fact that verbal and non-verbal attributes of teaching style affect student evaluations much more than the actual content not only questions the validity of SET per se, but also creates distorted incentives for the teachers. For instance, some professors seem to teach strategically to boost their SET scores, but harm the follow-on achievement of their students in more advanced classes (Carrell and West, 2010, Braga et al., 2014). Our study will reveal whether women fall prey of the relative importance of affect in student evaluations.

Third, one of our treatments proposes a solution to overcome gender bias in SET by making its effect on the teacher's payoff clear for the students. An average university student is unlikely to be aware of the incentive structure in academia. An explicit communication of the implications of SET scores for hiring and promotion decisions could make students evaluate their teachers more carefully. With growing awareness of the gender discrepancies in SET, universities start considering affirmative action policies to narrow the gap. For example, at the time of writing this paper, VU Amsterdam was considering to introduce a correction for the evaluation scores of female faculty based on the estimates of Mengel et al. (2019). We are aware of two experiments on the gender-related interventions into SET: Boring and Philippe (2017) and Peterson et al. (2019) nudged the students to be unbiased against women in their evaluations and observed a significant reduction of the gender gap. Our intervention is incentive compatible and much less suggestive, while still having the potential to induce the students to evaluate female teachers more objectively.

Finally, our study is related to the literature on attention to discrimination. The field experiment by Bartoš et al. (2016) was the first to show that the applications of minorities receive less attention on the job market and the rental market. Lahey and Oxley (2018) observed that employers spend less time on the resumes of young black applicants in a laboratory experiment with eyetracking. By measuring how much time the students spend on the SET questions in our experiment, we will be able to infer whether they invest less effort into the evaluation of female teachers and whether this channel contributes to gender bias in SET. This channel is related to attention discrimination, but measures selective deliberation on the experience with the teachers rather than selective information acquisition as in the two aforementioned experiments. We will control for information acquisition by the students by taking the amount of notes taken during the lecture as a proxy for their attention to the lecture.

The rest of the paper is organized as follows: Section 4.2 introduces the experimental design, Section 4.3 states the hypotheses and outlooks for data analysis, and Section 4.4 concludes and provides suggestions for further research. Instructions, the scripts of the lectures and the questionnaires can be found in Appendix D.

4.2 Experimental design

The participants of our experiment will act either as teachers or as students. The subjects in the role of the teachers will prepare a 20-minute video lecture for the students. The subjects in the role of the students will watch this videolecture, take an incentivized test based on its content and thereafter evaluate the teacher.

4.2.1 Timeline of the experiment

Phase 1: Teachers. In the first phase of our experiment, we will recruit 16 subjects for the role of teachers among the PhD students of University of Munich, excluding those at the chair for behavioral and experimental economics. Their task will be to give a 20-minute lecture on the numeral system of the ancient Maya (in English language). The topic is chosen to be related to mathematics since it is one of the fields where women are typically perceived to perform worse than men. We will ask the teachers to briefly introduce themselves, while being videotaped, and to read the script of the lecture prepared by us. The self-introduction video, and the audio record of the script along with the slides prepared by us will then be presented to the students to keep the non-verbal differences in teaching style minimal, following Chisadza et al. (2019). Depending on the treatments described below, the slides and the scripts will differ. We are going to keep the pace of reading the script the same across treatments. Each teacher is going to receive a fixed payment of 20 Euros and a chance to win the prize of 50 Euros, explained in detail in the next subsection.

Phase 2: Students. The second phase of the experiment will take place at the laboratory, with all subjects being in the role of the students. The laboratory experiment will consist of several parts. The instructions for each part will be given after the previous part has ended.

First, each subject will privately watch the prepared video lecture and take a test on its content. The test consists of 10 questions related to conversion of Mayan numerals into the decimal numeral system and simple algebraic operations with Mayan numbers. The time limit will be set to 20 minutes, and subjects will earn 1 Euro for each correct answer. We will provide the subjects with pens and paper for their notes during the lecture and the test.

After the test, we will ask the students to evaluate their teachers in a questionnaire. Different characteristics of teaching will be evaluated on a 5-point Likert scale. They include the overall impression of the lecture, the quality of the presentation, the motivation of the lecturer, etc. (see the questionnaire in Appendix D.3). Most questions are equivalent to those asked in SET at the University of Munich. The students will be aware that the evaluation is anonymous and the teachers are going to receive aggregated feedback scores. Since teaching in our experiment is primarily aimed at delivering information for the subsequent test (the so-called teaching-to-the-test), the evaluation takes place after the test is completed. Only thereafter it will be possible for the students to assess whether the lecture was adequate.

The evaluation will be followed by Raven's progressive matrices test, aimed at measuring the cognitive abilities of the subjects. The test consists of 8 tasks to identify a symbol that fits a given pattern. Each correct answer will be rewarded with 50 cents. Finally, a questionnaire on demographic characteristics and outcomes related to cognitive abilities (such as the grade for mathematics in high school), as well as additional evaluation questions about the lecture, will complete the experiment. The questionnaire can be found in Appendix D.4. The evaluation items on this questionnaire are related to the seductiveness of the teacher and long-term benefits of the lecture, discussed in detail in the next sections. These types of questions are not typical for university SET, but are of interest for our research, thereby we decided to ask them at the end of the experiment. It is explicitly stated that the teacher does not learn the answers to these questions, and they do not affect the evaluation of the teacher in the preceding part of the experiment.

We intend to conduct the experiment in the Research Lab at VU Amsterdam. The participants in the role of the students are going to be recruited from the subject pool of the laboratory. We will invite approximately the same number of female and male participants to create a balanced sample. The total number of subjects in the role of students will amount to 256 (statistical power is discussed in the next section). The subjects for the role of the teachers are intentionally recruited at a different university and a different level of "seniority" to avoid any information spillovers. We expect the experimental sessions to last about one hour and the average payment to be about 15 Euros.

4.2.2 Treatments

Our study is a 2×2×2 factorial design. Three dimensions of our treatment variation include the teacher's gender, the quality of teaching and the transparency of teacher's incentives, summarized in Table 4.1. Each subject in the role of a student will participate in only one treatment. There will be 32 subjects in each treatment condition (256 in total).

Table 4.1: Treatment variations in the experiment

		Transparency of incentives	
		PrizeInfo	*NoPrizeInfo*
Quality of teaching	*HighQ*	Male/Female	Male/Female
	LowQ	Male/Female	Male/Female

Teacher's gender. We randomly assign our students to either female or male lecturers. To minimize the differences besides gender, the teachers are going to be recruited among white non-native English-speakers of similar age. Courses at VU Amsterdam are frequently taught by PhD students in English language, so our sample of teachers will be to some degree similar to the ones students usually encounter. We intentionally keep the race fixed since it is also known to affect student evaluations, while our study focuses only on gender effects. A small photograph of the lecturer will be attached to all lecture slides to remind students of the teacher's gender. To make sure that potential treatment differences are driven not only by the appearance or accent of a particular teacher, we intend to recruit 8 male and 8 female subjects to perform as lecturers.

Quality of teaching. In order to exogenously vary the quality of teaching, we will present two versions of the slides and the script. In *HighQ* treatments, the lecture is designed to be helpful for the follow-on test, is well-structured and introduces many practical examples. The slides present all important information from the script and are animated to ease the visual perception of the mathematical content. In *LowQ* treatments, we intentionally degenerate the quality of the script by presenting the content in a messier way, leaving out some examples and adding chunks of irrelevant information. The slides are changed accordingly and the animation is left out. The complete scripts of the lecture along with the slides in both treatments can be found in Appendix D.2. To avoid experimenter demand effects, we will make each teacher participate in only one quality treatment condition. Thus, we will have 4 female and 4 male teachers in each of the two treatments.

Transparency of incentives. As mentioned above, the teachers have the possibility to win a prize of 50 Euros. This prize is given to the teacher who receives the highest average score[15] according to the student evaluations in our experiment. Our last treatment variation concerns the transparency of this prize to the students. In *NoPrizeInfo* treatment, the students are unaware of how the teachers are paid. In *PrizeInfo* treatment, the experimenter communicates the existence of the prize to the students before they start with their evaluations. In either of the treatments, we will show the same records of the teachers (either high or low quality). To make the chances of winning the prize equal, we will give two separate prizes for *HighQ* and *LowQ* teachers and guarantee the same number of evaluations for each teacher. Since there are 4 teachers and 32 students in each treatment condition, and the records of the same teachers are shown in *PrizeInfo* and *NoPrizeInfo* treatments, the record of each teacher will be shown to 16 students.

The instructions for the experiment are presented in Appendix D.1.

[15]The average score is taken over all evaluation questions and over all treatments, in which the respective teacher's records were shown.

4.3 Hypotheses and outlooks for data analysis

This section elaborates on our expected results and the methods of empirical analysis. To test Hypotheses 4.1, 4.2 and 4.3, we are going to analyze the following outcome variables (refer to the questionnaire in Appendix D.3):

- the average score on all evaluation items;
- subjective grade for the lecture;
- subjective grade for the teacher.

Hypothesis 4.1. Female teachers receive lower evaluation scores than male teachers in all treatment conditions.

We intend to replicate the gender bias prevalent in previous studies on SET. Although the experiment by Chisadza et al. (2019) found a reverse effect, we expect our results to be in line with the other literature since we select a male-dominated discipline and add more variance both to the pool of the lecturers and to the pool of the students. We are going to estimate the gender bias in each of the four treatment cells separately: *HighQ* and *LowQ* treatment conditions combined with *PrizeInfo* and *NoPrizeInfo* treatment conditions. The bias might shrink or disappear in some of the treatments (see further hypotheses below), but on average we expect the evaluations of female lecturers to be at least weakly worse than those of their male counterparts. We expect this to hold for all three outcome variables mentioned above. Even though the subjective grade for the lecture is not directly related to the teacher, previous studies have found that female teachers fare worse scores also on the evaluation items not related to the instructor (see Mengel et al., 2019). We will apply pairwise Mann-Whitney U tests[16] in each treatment cell to each of the three outcome variables of female versus male teachers.

Hypothesis 4.2. Poorer quality of the lecture affects evaluation scores negatively, with this effect being stronger for female teachers.

On the one hand, the variation of the lecture content serves as a manipulation check: naturally, lower quality of teaching should result in poorer evaluations, at least on the

[16] As a robustness check, we are going to use the Epps-Singleton test: unlike the MWU test, it is able to detect non-locational shifts in the distribution.

lecture-related items. On the other hand, relying on the results of psychological experiments, we expect this relationship to be stronger for women than for men. Ware and Williams (1975) showed that highly seductive (i.e. confident, humorous, gesticulating) teachers receive high evaluations irrespective of the content they deliver, while content mattered only under low seductiveness. Although the differences in teaching are minimized in our experiment, we hypothesise that male teachers are implicitly associated with a more seductive teaching style as female teachers. In this case, female lecturers would be punished more in SET for poorer lecture quality. We will be able to control for the perceived seductiveness of the teacher by including additional items evaluating different dimensions of teacher's seductiveness into the questionnaire at the end of the entire experiment (refer to Appendix D.4).

To estimate the treatment effect of quality, we are going to compare the same three evaluation measures as above in *HighQ* versus *LowQ* treatments using Mann-Whitney U test, and then compare the size of the treatment effect for women and men. Naturally, we expect a difference in the overall average score and the subjective grade for the lecture; whether the subjective grade for the teacher is also affected by variations in lecture quality, remains an empirical question.

Hypothesis 4.3. The communication of teachers' incentives lowers the gender bias in evaluation scores.

By announcing in *PrizeInfo* treatments that the teachers are competing for a monetary reward based on SET, we appeal to students' fairness and expect them to invest more cognitive resources into the evaluation of both female and male teachers. We speculate that one potential channel of gender bias in SET is more emotional or less careful evaluation of female teachers. For instance, one semantic analysis of a large mass of online evaluations revealed that female professors receive more comments on their appearance and personality and are referred to less respectfully (Arceo-Gomez and Campos-Vazquez, 2019). Our intervention might push the students to invest more effort into the questionnaires, potentially leading to more objective grades and less gender bias. Note that this not the only possible channel that could be activated by our intervention: e.g., the knowledge of the prize might induce the students to reciprocate and give higher evaluations to the teachers they liked the most, and consequently exacerbate the gender bias. However, we expect fairness and objectiveness to be the dominant channel in our setting.

We will compare the difference in female and male evaluation scores between *PrizeInfo* and *NoPrizeInfo* treatments using Mann-Whitney U test. The three outcome measures stay the same as in Hypotheses 4.1 and 4.2. Most of all we are interested in the improvements in the subjective grade for the teacher and the overall average score.

Hypothesis 4.4. Students spend less time on the evaluations of female teachers compared to male teachers, this difference being less pronounce in *PrizeInfo* treatments. They also pay less attention to the lectures given by female teachers in comparison to male teachers.

In line with the findings of Bartoš et al. (2016) and Lahey and Oxley (2018), we expect that the students pay less attention both to the lecture and the evaluation of female teachers. As discussed above, inattention to evaluation, or selective deliberation, might be mitigated by the intervention in *PrizeInfo* treatment. To assess attention to evaluation, pairwise Mann-Whitney U tests will be applied to the time that the students spend on the screen with the evaluation of their female or male teachers (separately in each treatment condition). We will then compare the treatment effect of teacher's gender across *PrizeInfo* and *NoPrizeInfo* treatments. Attention to the lecture will be compared using Fisher's exact test on a binary variable indicating that a student took notes during the lecture (in each treatment condition).

Statistical power of Mann-Whitney U tests was calculated using the simulation method proposed by Campos-Mercade (2018). We took the data from Chisadza et al. (2019) as a baseline: the authors found that male teachers received an average score of 3.745 points on a 5-point Likert scale with a standard deviation of 0.0419. Since the treatment differences in that experiment went in an unexpected direction (probably because of experimental design features we discussed in the introduction), we refer to the estimates of Mengel et al. (2019) to calculate our expected treatment effect. The authors estimated that male students evaluated female instructors 20.7% of a standard deviation worse than male instructors, and female students did so by 7.6%. It translated into 0.2 points and 0.07 points on Likert scale, respectively. Since our student sample will be gender-balanced, the average treatment effect is expected to be roughly 0.13 Likert scale points. This is also consistent with the treatment effects found by Chisadza et al. (2019), though their differences had an opposite sign. Combined with the expected baseline evaluations, a small sample of 5 students per treatment condition would thus be enough to achieve the statistical power of 98%. To estimate the effect of quality treatments, we refer to the study by Ware and Williams (1975). They found that low seductiveness combined with poorer quality of content resulted in a drop of evaluation scores from 48.7 to 35.5 points on a different scale, or a difference of 86.6% of a standard deviation. Applied to the baseline standard deviation of 0.0419 found by Chisadza et al. (2019), we derive an expected treatment effect of 0.036 points on the Likert scale. Sample size of 32 students per treatment would yield statistical power of roughly 91%. We have no expectation to what extent our *PrizeInfo* treatment will influence evaluation scores since this kind of intervention in SET was not previously studied. Thus, we decided to have 32 subjects in each treatment condition. A $2 \times 2 \times 2$ design implies 8 treatment cells, resulting in a sample of 256 subjects in total.

Apart from non-parametric comparison tests, we want to prove the robustness of our results by regression analysis. We will estimate two following OLS regression models with robust standard errors clustered on teacher level:

$$y_i = \beta_1 female + \beta_2 LowQ + \beta_3 female \times LowQ + \beta_4 X_i + \epsilon_i; \tag{3}$$

$$y_i = \beta_1 female + \beta_2 PrizeInfo + \beta_3 female \times PrizeInfo + \beta_4 X_i + \epsilon_i. \tag{4}$$

The outcome variables encompass three evaluation scores used to test Hypotheses 4.1, 4.2 and 4.3, as well as the time spent on the questionnaire (Hypothesis 4.4). Variable $female$ is the dummy of the lecturer's gender, and variables $LowQ$ and $PrizeInfo$ are the respective treatment dummies. The vector of control variables X_i includes the gender of the student, his academic major, his grade for mathematics in high school and the score on the Raven's test. We intentionally do not include the student performance on the test in the experiment since it might be influenced by our treatments. Among the control variables, we expect to find significant effects of the gender of the student on the evaluations: previous literature has shown that male students rate female professors worse than female students do (see e.g. Chisadza et al., 2019, Mengel et al., 2019). We will run an additional specification of the two models with the interaction between the student's and the teacher's gender. Apart from that, the measures of cognitive ability are likely to be positively correlated to the performance on the test, which, in turn, might be correlated with the evaluation scores.

Both models 3 and 4 will be estimated on pooled data from all treatments; the choice of the models was motivated by the easiness of interpretation of the coefficients. For the sake of completeness, we are going to run an OLS regression on pooled data with all possible interactions between the treatment dummies and teacher's gender (the reference group is male teachers in *HighQ NoPrizeInfo* treatment):

$$
\begin{aligned}
y_i = {} & \beta_1 female \times LowQ \times PrizeInfo + \beta_2 female \times HighQ \times PrizeInfo + \\
& + \beta_3 female \times LowQ \times NoPrizeInfo + \beta_4 female \times HighQ \times NoPrizeInfo + \\
& + \beta_5 male \times LowQ \times PrizeInfo + \beta_6 male \times HighQ \times PrizeInfo + \\
& + \beta_7 male \times LowQ \times NoPrizeInfo + \beta_8 X_i + \epsilon_i.
\end{aligned} \tag{5}
$$

As a robustness check, we are going to estimate models 3, 4 and 5 as tobit regressions when the outcome variable is the average score over all evaluation items; and ordered logit regressions when the outcome variables are subjective grades for the lecture or the teacher.

4.4 Discussion and conclusion

We proposed an experimental design to investigate the determinants of gender bias in student evaluations of teaching in a controlled laboratory environment. We exogenously vary the gender of the teacher and the quality of the teaching material. Additionally, we designed a non-invasive intervention to mitigate the gender bias in SET (if any) by explaining the incentives of the teachers to the students. By measuring the time spent on the evaluation forms, we can detect how much of the gender bias is attributed to selective deliberation.

Although the focus of our study has been on the context of academia, its results could have potential implications for evaluation procedures outside the universities. For instance, some companies regularly ask their employers to evaluate their team leaders, who later receive this feedback. Feedback systems are widely used in many online environments (for an example of gender-based discrimination on an online platform, see Bohren et al., 2019).

One caveat of our experimental design is that we trade some features of the natural classroom environment for the control over the treatment variations in the laboratory. For instance, our experiment implies no interaction between students and teachers, which is an important part of student-teacher relationship in the field. Therefore, students will certainly feel less attached to the teachers in the laboratory; however, it is not clear in which direction it will drive the gender bias. It is also not common for the students in the lab to evaluate the lecturers, so we can expect some noise in their responses. To make the decision more relevant for the students, we announce that the teachers will learn their average evaluation scores – just like they do in the field. The evaluation procedure will unlikely be perceived as the central part of the experiment, so we expect the student responses to be a good measure of their real perceptions of the teachers.

There are many open questions around the gender bias in SET not covered by our study. First, there is a whole universe of potential affirmative action policies that can be tested both in the lab and in the field. For example, the upward correction of the SET scores for female faculty proposed by some universities is a double-edged sword: on the one hand, it reduces the gender gap in SET, and on the other hand, it is not incentive compatible. Besides, if the students are aware of this policy, they might evaluate women even worse.

Second, in our study we fully focus on the behavior of students. However, the behavior of teachers in presence of gender bias in SET deserves particular attention. To our knowledge, there is only observational evidence (Carrell and West, 2010, Braga et al., 2014) and no experimental studies that identify strategic teaching aimed at increasing the evaluation scores. This behavior might be even more prevalent in presence of real teaching prizes: just like we gave a prize for the teacher with the best evaluations in our experiment, university teachers also frequently compete for financial and non-financial teaching awards. The reaction of the teacher to these incentives might depend on the gender. The response of teachers to affirmative action is also of crucial importance.

Third, apart from gender, there many other dimensions that potentially influence subjective evaluations, such as race, age, look, etc. Race and ethnicity have been found to affect SET both in the lab and in the field (Chisadza et al., 2019, Reid, 2010, Wang and Gonzalez, 2020). It would be interesting to investigate how these factors interplay between teaching quality and the discipline.

Last but not least, could gender bias in SET result from asking the wrong questions? By design, our experiment was primarily concerned with teaching-to-the-test and so was the evaluation procedure. However, the goals of teaching in the field are much broader than preparing the students for an exam. For example, improving students' skills, influencing their way of thinking, helping them approach problems in life and building bridges between other disciplines are among them. Nevertheless, the evaluation questionnaires used at the universities are largely evaluating teaching-to-the-test rather than "teaching-for-life". We are not aware of any study trying to make the evaluation questions reflect these important aspects of teaching. One could imagine that women are more intrinsically motivated to contribute to the students' knowledge rather than achieve high evaluation scores in the current system, which does not reflect this contribution to the full extent. We included some of the items evaluating the global benefits of teaching into the questionnaire in Appendix D.4; although the design of the evaluation procedure is not the focus of our study, we will at least be able to see whether the gender bias is present on those dimensions in our simple setting. We also speculate that the results of any evaluation would be different if it was conducted several years after the course rather than directly after it. An evaluation delayed in time would make it possible to assess the long-term benefits of teaching. The intertemporal effects of student evaluations of teaching and their interaction with gender bias are left for future research.

Appendix A Two Heads are Better than One: Teams and Individuals in Standard Auction Formats

A.1 Regression tables

Table A.1: Profit in first-price auctions

	(1)	(2)	(3)	(4)
Individual	-2.048**	-3.486**	-2.937**	-2.559*
	(0.875)	(1.509)	(1.303)	(1.345)
Period		-0.0717	-0.0717	-0.0717
		(0.0467)	(0.0468)	(0.0469)
Individual*Period		0.221**	0.221**	0.221**
		(0.110)	(0.111)	(0.111)
SPA error			-0.171***	-0.168***
			(0.0580)	(0.0558)
Risk			0.0255*	0.0181*
			(0.0150)	(0.0110)
Cognitive			0.270	0.398
			(0.432)	(0.248)
Female				-2.437***
				(0.902)
Math grade				0.417
				(0.647)
Experience				0.389
				(0.578)
Constant	3.953***	4.419***	2.110	0.676
	(0.197)	(0.297)	(2.547)	(3.449)
Observations	792	792	792	792
R^2	0.0152	0.0177	0.0892	0.106

GLS random effects regression. Robust standard errors in parentheses (clustered on matching group level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Bids in first-price auctions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual	2.663*	2.719**	2.848**	2.164	1.237	4.584	3.693	2.781
	(1.573)	(1.382)	(1.364)	(1.439)	(1.316)	(3.453)	(3.385)	(3.255)
Value	0.824***	0.824***	0.826***	0.828***	0.828***	0.825***	0.824***	0.824***
	(0.0137)	(0.0135)	(0.0127)	(0.0126)	(0.0126)	(0.0136)	(0.0138)	(0.0139)
Individual*Value		-0.00117	-0.00384	-0.00819	-0.00780			
		(0.0260)	(0.0258)	(0.0260)	(0.0264)			
Period			0.242	0.242	0.242	0.402***	0.402***	0.402***
			(0.185)	(0.185)	(0.186)	(0.123)	(0.123)	(0.124)
SPA error				0.130***	0.160***		0.130***	0.159***
				(0.0476)	(0.0531)		(0.0473)	(0.0529)
Risk				-0.0497*	-0.0575*		-0.0497*	-0.0575*
				(0.0281)	(0.0317)		(0.0280)	(0.0316)
Cognitive				-1.600**	-2.252***		-1.590**	-2.240***
				(0.653)	(0.704)		(0.658)	(0.696)
Female					0.339			0.357
					(2.220)			(2.208)
Math grade					-2.417*			-2.410*
					(1.371)			(1.371)
Experience					-2.418***			-2.426***
					(0.733)			(0.729)
Individual*Period						-0.296	-0.296	-0.295
						(0.337)	(0.338)	(0.338)
Constant	-1.020	-1.051**	-2.700**	9.291*	27.33***	-3.682***	8.369*	26.38**
	(0.705)	(0.418)	(1.321)	(4.997)	(10.03)	(1.022)	(4.831)	(10.25)
Observations	792	792	792	792	792	792	792	792
R^2	0.836	0.836	0.837	0.852	0.860	0.838	0.852	0.861

GLS random effects regression. Robust standard errors in parentheses (clustered on matching group level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Prices in first-price auctions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual	4.842	5.422	6.717	4.613	1.743	0.191	-1.576
	(3.521)	(3.539)	(4.376)	(4.111)	(4.438)	(4.315)	(4.238)
Value	0.831***	0.838***	0.870***	0.855***	0.828***	0.831***	0.820***
	(0.0548)	(0.0536)	(0.0595)	(0.0583)	(0.0392)	(0.0412)	(0.0400)
Individual*Value	0.00120	-0.0108	-0.0709	-0.0628			
	(0.0774)	(0.0787)	(0.0737)	(0.0780)			
Period		0.380	0.391	0.389	0.113	0.114	0.109
		(0.292)	(0.289)	(0.290)	(0.156)	(0.156)	(0.157)
SPA error			0.237	0.249**		0.232	0.242**
			(0.146)	(0.116)		(0.148)	(0.119)
Risk			-0.102*	-0.108**		-0.100*	-0.106**
			(0.0576)	(0.0462)		(0.0570)	(0.0457)
Cognitive			-2.872*	-4.358***		-2.791*	-4.279***
			(1.672)	(1.568)		(1.683)	(1.578)
Female				0.684			0.677
				(4.882)			(4.864)
Math grade				-4.427***			-4.470***
				(1.594)			(1.575)
Experience				-4.704*			-4.745*
				(2.676)			(2.672)
Individual*Period					0.486	0.484	0.492
					(0.524)	(0.525)	(0.521)
Constant	19.76***	16.97***	37.74***	73.94***	19.15***	40.84***	77.11***
	(2.982)	(3.343)	(7.135)	(16.71)	(1.746)	(7.634)	(15.56)
Observations	264	264	264	264	264	264	264
R^2	0.481	0.486	0.517	0.530	0.487	0.518	0.531

GLS random effects regression, data collapsed on market level. Robust standard errors in parentheses (clustered on matching group level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Profit in English auctions

	(1)	(2)	(3)	(4)
Individual	-0.672	-3.885	-3.590	-4.134
	(1.243)	(2.399)	(2.662)	(2.700)
Period		-0.304	-0.304	-0.304
		(0.248)	(0.248)	(0.249)
Individual*Period		0.714*	0.714*	0.714*
		(0.380)	(0.382)	(0.383)
SPA error			-0.0532***	-0.0677***
			(0.0142)	(0.0215)
Risk			0.00351	0.00143
			(0.0130)	(0.0232)
Cognitive			1.027	1.104*
			(0.728)	(0.659)
Female				0.447
				(2.317)
Math grade				1.498**
				(0.671)
Experience				-0.0667
				(1.284)
Constant	6.904***	8.274***	2.059	-1.333
	(0.656)	(1.565)	(5.432)	(7.501)
Observations	528	528	528	528
R^2	0.000511	0.00372	0.0148	0.0208

GLS random effects regression. Robust standard errors in parentheses (clustered on matching group level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Prices in English auctions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual	-7.120	-7.226	-6.389	-6.843	15.98*	16.19*	15.43*
	(7.853)	(8.024)	(7.890)	(7.777)	(8.495)	(8.555)	(8.840)
Value	0.772***	0.780***	0.786***	0.781***	0.875***	0.875***	0.867***
	(0.155)	(0.151)	(0.155)	(0.160)	(0.0839)	(0.0830)	(0.0897)
Individual*Value	0.180	0.182	0.170	0.166			
	(0.170)	(0.175)	(0.182)	(0.180)			
Period		0.712	0.712	0.706	2.394**	2.393**	2.389**
		(0.745)	(0.753)	(0.759)	(1.103)	(1.114)	(1.123)
SPA error			-0.118	-0.137		-0.135	-0.152
			(0.148)	(0.155)		(0.145)	(0.152)
Risk			0.000663	0.0197		-0.00280	0.0158
			(0.0857)	(0.0929)		(0.0838)	(0.0897)
Cognitive			0.193	0.562		0.0950	0.491
			(1.383)	(1.707)		(1.302)	(1.677)
Female				5.495			5.706
				(9.408)			(9.634)
Math grade				1.336			1.330
				(4.206)			(4.112)
Experience				0.0544			-0.206
				(3.411)			(3.238)
Individual*Period					-3.099**	-3.099**	-3.101**
					(1.244)	(1.255)	(1.264)
Constant	10.02	6.413	5.713	-3.047	-6.110	-5.499	-13.41
	(6.587)	(6.736)	(7.511)	(23.46)	(6.163)	(9.997)	(26.19)
Observations	176	176	176	176	176	176	176
R^2	0.432	0.437	0.444	0.444	0.459	0.467	0.467

GLS random effects regression, data collapsed on market level. Robust standard errors in parentheses (clustered on matching group level). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Emotions associated with winning an auction

	Dependent variable			
	Neutrality	Happiness	Sadness	Anger
Individual	-0.0351	-0.0272	0.0144**	0.00985
	(0.0271)	(0.0176)	(0.00709)	(0.0125)
Winner	-0.00843	0.0112**	-0.00479	-0.00317
	(0.00609)	(0.00468)	(0.00605)	(0.00408)
Individual*Winner	-0.00805	-0.00656	-0.000756	-0.0120**
	(0.0242)	(0.0107)	(0.0112)	(0.00554)
Constant	0.650***	0.105***	0.0795***	0.0600***
	(0.0119)	(0.00678)	(0.00300)	(0.00863)
Observations	1500	1500	1500	1500
R^2	0.00519	0.00765	0.00508	0.00102

	Dependent variable			
	Surprise	Fear	Disgust	Valence
Individual	0.00633	-0.000877	0.00546	-0.0471**
	(0.00543)	(0.000807)	(0.00511)	(0.0215)
Winner	-0.00415*	0.000377	0.000776	0.0177***
	(0.00248)	(0.00135)	(0.00252)	(0.00475)
Individual*Winner	0.00505	-0.000942	0.00661	0.000616
	(0.00523)	(0.00137)	(0.00492)	(0.0175)
Constant	0.0213***	0.00170**	0.00755**	-0.0225**
	(0.00417)	(0.000762)	(0.00300)	(0.0109)
Observations	1500	1500	1500	1500
R^2	0.00485	0.00123	0.0111	0.0103

GLS random effects regression. Robust standard errors in parentheses (clustered on matching group level).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: The effect of emotions on future bids

	(1)	(2)	(3)	(4)	(5)
Individual	2.819	2.345	2.879	2.842	2.209
	(1.931)	(1.810)	(1.836)	(1.840)	(1.663)
Value	0.834***	0.835***	0.835***	0.833***	0.834***
	(0.0135)	(0.0136)	(0.0131)	(0.0105)	(0.0106)
Individual*Value	-0.00240	-0.00561	-0.00346		
	(0.0222)	(0.0225)	(0.0211)		
Period	0.0421	0.0420	0.0395	0.0396	0.0396
	(0.166)	(0.166)	(0.165)	(0.163)	(0.163)
Valence_{-1}			1.010	0.361	0.403
			(1.327)	(0.639)	(0.687)
$\text{Individual*Valence}_{-1}$				2.430	2.535
				(3.822)	(3.723)
Neutral_{-1}	0.797	0.776			
	(1.093)	(1.151)			
Happy_{-1}	1.375	1.657			
	(1.930)	(2.040)			
Sad_{-1}	-4.246**	-3.859**			
	(1.686)	(1.668)			
Angry_{-1}	2.783	2.763			
	(2.003)	(2.022)			
Surprised_{-1}	11.36	11.87			
	(10.93)	(10.80)			
Scared_{-1}	18.67***	18.01***			
	(4.736)	(4.585)			
Disgusted_{-1}	-1.078	-2.112			
	(3.806)	(3.773)			
Constant	-1.898	6.479	-1.152	-1.104	7.070
	(1.468)	(4.780)	(1.274)	(1.228)	(5.212)
Controls	No	Yes	No	No	Yes
Observations	1375	1375	1375	1375	1375
R^2	0.902	0.908	0.902	0.902	0.907

GLS random effects regression. Robust standard errors in parentheses (clustered on matching group level).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.2 Instructions

Below we provide the instructions in *TEAM FPA* treatment (in German).

Willkommen beim Experiment und vielen Dank für Ihre Teilnahme!

Bitte sprechen Sie von nun an nicht mit anderen Teilnehmern des Experiments.

Allgemeines zum Ablauf

Dieses Experiment dient der Untersuchung von ökonomischem Entscheidungsverhalten. Sie können dabei Geld verdienen. Dieses wird Ihnen im Anschluss an das Experiment privat und in bar ausbezahlt.

Das gesamte Experiment dauert etwa 2 Stunden und besteht aus 4 Teilen. Zu Beginn jeden Teils erhalten Sie detaillierte Instruktionen. Die Teile sind unabhängig voneinander; Entscheidungen in einem Teil haben keine Auswirkungen auf Ihren Verdienst in einem anderen Teil. Die Summe Ihrer Einkommen aus den 4 Teilen ergibt Ihren Gesamtverdienst aus dem Experiment. Dieser wird Ihnen nach Abschluss des vierten Teils mitgeteilt und am Ende des Experiments in bar ausbezahlt.

Wenn Sie nach den Instruktionen oder während des Experiments Fragen haben, heben Sie bitte Ihre Hand. Einer der Experimentleiter wird dann zu Ihnen kommen und Ihre Fragen privat beantworten. Der sprachlichen Einfachheit halber verwenden wir nur männliche Bezeichnungen.

Während des Experiments werden Sie bzw. die anderen Teilnehmer gebeten, Entscheidungen zu treffen. Sie werden dabei zum Teil mit anderen Teilnehmern interagieren. Das heißt, sowohl Ihre eigenen Entscheidungen als auch die Entscheidungen der anderen Teilnehmer können Ihre Auszahlung bestimmen. Diese ergibt sich entsprechend der Regeln, die im Folgenden erklärt werden.

Während Sie Ihre Entscheidungen treffen, wird am rechten oberen Rand des Bildschirms eine Uhr herunterlaufen. Diese bietet Ihnen eine Orientierung, wie viel Zeit Sie für Ihre Entscheidung benötigen. Manchmal ist diese Zeitrestriktion bindend, manchmal nicht. Details werden weiter unten erläutert. Die Informationsbildschirme, bei denen keine Entscheidungen zu treffen sind, werden nach Ablauf der Uhr ausgeblendet.

Bezahlung

In den Teilen 1-3 des Experiments sprechen wir nicht von Euro, sondern von Punkten. Diese werden am Ende des Experiments in Euro umgerechnet. Der Wechselkurs für die Umrechnung wird am Anfang jedes Teils bekanntgegeben. In Teil 4 sprechen wir von Euro.

Für Ihr pünktliches Erscheinen erhalten Sie 4 € zusätzlich zu dem Einkommen, das Sie während des Experiments verdienen können.

Anonymität

Wir werten die Daten aus dem Experiment nur im Aggregat aus und verknüpfen Namen nie mit den Daten aus den Experimenten. Am Ende des Experiments müssen Sie eine Quittung über den Erhalt des Verdienstes unterschreiben, die nur der Abrechnung dient und keine Verknüpfung mit Ihrem Verhalten im Experiment zulässt. Während des Experiments werden Videoaufnahmen angefertigt. Diese dienen dem Forschungszweck des Experiments. Verhalten Sie sich einfach ganz natürlich.

Hilfsmittel

An Ihrem Platz finden Sie einen Kugelschreiber. Bitte lassen Sie ihn nach dem Experiment am Tisch liegen.

Gruppen

Im Experiment werden Sie ein Mitglied in einer von sechs Gruppen sein. Jede Gruppe besteht aus drei Mitgliedern. Die Zuteilung in die Gruppen erfolgt zufällig. Die Gruppen bleiben über das gesamte Experiment hinweg unverändert.

Teil 1

Wechselkurs

Der Wechselkurs in Teil 1 ist: **3 Punkte = 1 Euro**

Auktionen

Teil 1 besteht aus zwölf unabhängigen Auktionen. In jeder Auktion bieten Sie in Ihrer Gruppe für ein Gut, das einen bestimmten Wert für Sie bzw. Ihre Gruppe hat. Drei Gruppen werden für jede Auktion zufällig verbunden, um für ein Gut zu bieten. Wir nennen das einen Markt (bestehend aus drei Gruppen á drei Mitglieder).

Jedes Gruppenmitglied erhält eine einmalige Anfangsausstattung von 10 Punkten.

Wert des Gutes

Jede Gruppe erhält zu Beginn jeder Auktion die Information über den Wert des Gutes. Dieser Wert wird für jede Gruppe auf einem Markt unabhängig bestimmt; d.h., es ist sehr wahrscheinlich, dass er in derselben Auktion für verschiedene Gruppen unterschiedlich ist. Der Wert für jede Gruppe wird zufällig zwischen 0 und 100 Punkten gezogen. Dabei ist jeder Wert zwischen 0 und 100 gleichwahrscheinlich (wobei 0 und 100 auch möglich sind). Jede Bietergruppe kennt ihren eigenen Wert, aber nicht die Werte der beiden anderen Gruppen auf dem Markt.

Entscheidungsfindung

In diesem Teil wird jede Gruppe ein Gebot abgeben. Jede Gruppe entscheidet sich gleichzeitig für ein Gebot. Die Entscheidung innerhalb der Gruppe erfolgt in zwei Stufen:

In der ersten Stufe werden Sie individuell um einen Vorschlag für das Gruppengebot gefragt. Sie haben 30 Sekunden Zeit, einen Vorschlag in der Mitte des Bildschirms einzugeben und auf "Weiter" zu klicken. Nachdem alle Gruppenmitglieder Ihrer Gruppe den Vorschlag eingegeben haben, startet die zweite Stufe.

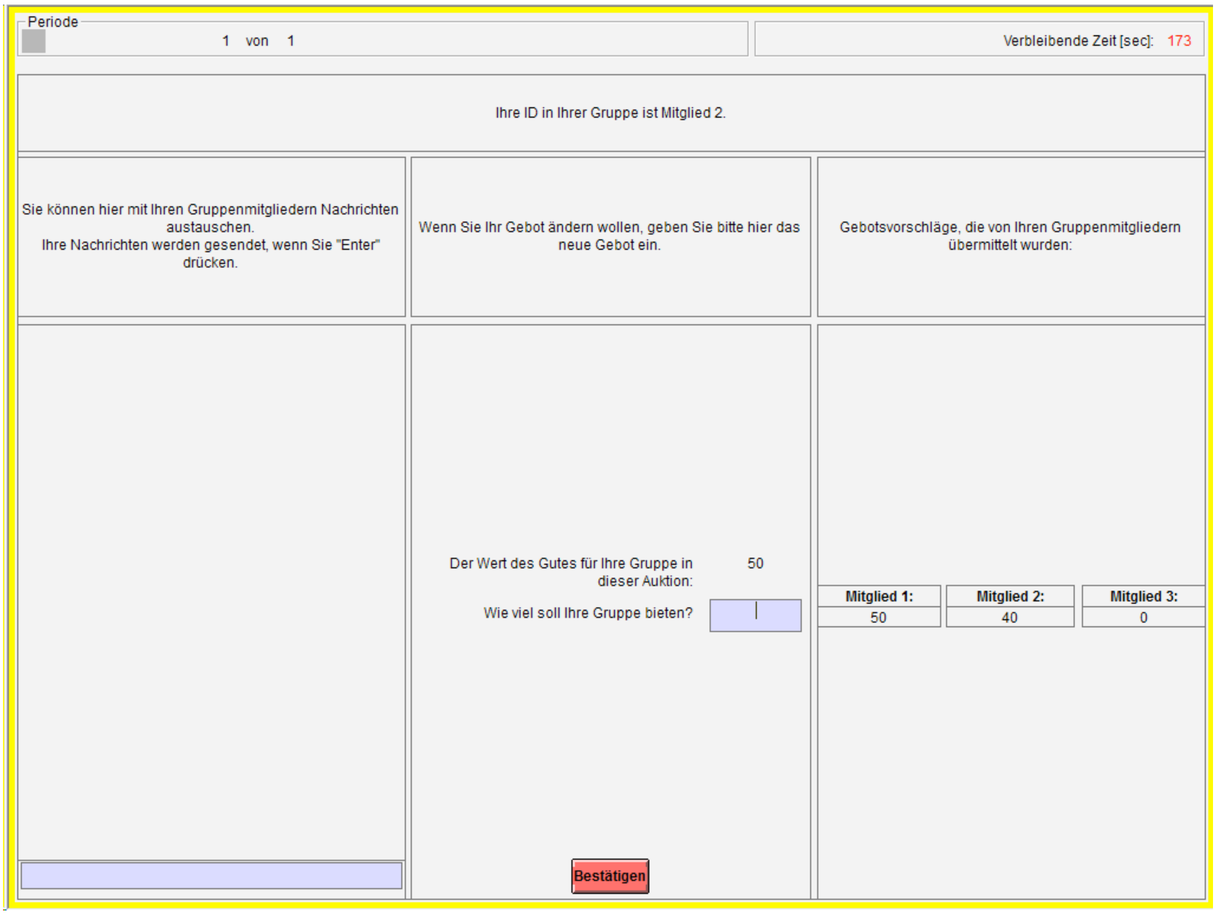
In der zweiten Stufe hat die Gruppe 180 Sekunden, um sich auf eine gemeinsame Entscheidung zu verständigen. Eine Entscheidung ist nur dann gültig, wenn von den drei Gruppenmitgliedern jeweils exakt dieselbe Entscheidung eingegeben wird (Entscheidungen Ihrer Gruppenmitglieder sehen Sie im rechten Teil Ihres Bildschirms). Innerhalb der 180 Sekunden können Sie und die anderen Gruppenmitglieder ihre Entscheidungen so oft anpassen, wie Sie wollen. Sie können Ihre Entscheidungen mit den anderen Gruppenmitgliedern auch diskutieren – diese Diskussion erfolgt über ein Chat-Fenster im linken Teil des Bildschirms.

Für den Chat gelten folgende Regeln: (i) deutschsprachige Konversation; (ii) keine Beleidigungen, Drohungen oder ähnliche Verstöße gegen die Etikette; (iii) keine Informationen, die es erlauben würden Sie zu identifizieren (z.B. Sitzplatznummer, Namen, Geschlecht, Studienrichtung etc.). Wenn Sie gegen diese Regeln verstoßen, werden Sie aus dem Experiment ausgeschlossen.

Der Chat soll Ihnen helfen, sich innerhalb der Gruppe zu koordinieren. Wenn es eine Gruppe nicht schafft, innerhalb von 180 Sekunden eine gemeinsame Entscheidung zu treffen, wird ein Gebot von 0 angenommen.

Gebote müssen immer ganzzahlig sein. Das Höchstgebot beträgt 110 Punkte.

Alle Informationen finden Sie dann auch auf dem Bildschirm, der wie folgt aussehen wird:



Im oberen Bereich des Bildschirms sehen Sie Ihre Identifikationsnummer (ID) in Ihrer Gruppe (z.B."Ihre ID in Ihrer Gruppe ist Mitglied 2"). Ihre ID wird für das gesamte Experiment beibehalten.

Auf der linken Seite des Bildschirms sehen Sie ein Chat-Fenster, in dem Sie mit den anderen Teilnehmern Ihrer Gruppe Nachrichten austauschen können. Geben Sie hierzu in das violette Feld unten links Ihre Nachricht ein und drücken Sie dann die Eingabetaste (Return/Enter). Ihre Eingabe wird dann an den Computer übermittelt und im grauen Bereich des Fensters erscheinen. Die anderen Teilnehmer Ihrer Gruppe sehen dann Ihre Nachricht und auch Sie sehen die Nachrichten der anderen Teilnehmer Ihrer Gruppe oberhalb des violetten Eingabe-Feldes. Wichtig: Wenn Sie eine Nachricht senden möchten, drücken Sie in jedem Fall Eingabetaste (Return/Enter), so dass der von Ihnen geschriebene Text im grauen Bereich erscheint.

Auf der rechten Seite des Bildschirms sehen Sie die Vorschläge der Mitglieder Ihrer Gruppe für das Gebot. Wenn Sie sich dafür entscheiden, Ihr Gebot zu ändern, können Sie dies in der Mitte des Bildschirms tun. Ihr Gebot wird geändert, wenn Sie ein neues Gebot in das Eingabefeld in der Mitte eingeben und auf "Bestätigen" klicken. Ihr neues Gebot wird dann rechts auf dem Bildschirm Ihr altes Gebot ersetzen.

Auszahlungen

Die Gruppe mit dem höchsten Gebot der drei Gruppen gewinnt die Auktion. Sie bezahlt für das Gut als Preis ihr Gebot. Jedes Gruppenmitglied verdient also in dieser Auktion:

Auszahlung jedes Gruppenmitglieds = Wert des Gutes – das Gebot seiner Gruppe

Zum Beispiel: Die drei Gruppengebote sind 80, 60, 40; die Gruppe, die 80 Punkte geboten hat, gewinnt die Auktion. Jedes Mitglied dieser Gruppe erhält den Wert des Gutes und bezahlt 80 Punkte.

Wenn alle drei Gruppen 0 geboten haben, wird das Gut nicht verkauft. Wenn zwei oder drei Bietergruppen das gleiche, höchste Gebot (größer 0) abgegeben haben, wird zufällig bestimmt, welche Gruppe das Gut erhält und bezahlt.

Wenn eine Gruppe das Gut in einer Auktion nicht erhält, verdient diese Gruppe in dieser Auktion nichts.

Jedes Gruppenmitglied behält allerdings die Anfangsausstattung von 10 Punkten (mit einer Ausnahme: siehe weiter unten).

Feedback und weitere Auktionen

Ist eine Auktion beendet, startet eine neue Auktion mit exakt den gleichen Regeln. Es gibt insgesamt 12 Auktionen in diesem Teil des Experiments. In jeder Auktion werden die sechs Bietergruppen im Raum wieder zufällig neu zu zwei Märkten zusammengewürfelt. Die drei Mitglieder einer Gruppe bleiben dabei aber unverändert.

Am Ende des Experiments wird eine Auktion zufällig als auszahlungsrelevant ausgelost. Ihre Auszahlung aus Teil 1 ergibt sich aus den Punkten, die Sie in dieser Auktion verdient haben (also: Wert minus Gebot, wenn Ihre Gruppe die Auktion gewonnen hat; oder null, wenn Ihre Gruppe die Auktion nicht gewonnen hat) plus die Anfangsausstattung von 10 Punkten.

Achtung!

Sie können auch Verluste in einer Auktion machen. Wenn Ihre Gruppe mit einem Gebot gewinnt, das über dem Wert des Gutes liegt, dann machen Sie Verluste! Verluste werden von Ihrer Anfangsausstattung abgezogen (und wenn diese nicht reicht: von Gewinnen aus anderen Teilen des Experiments).

Wenn Sie Fragen haben, dann heben Sie bitte Ihre Hand. Ein Experimentleiter kommt dann zu Ihnen, um Ihre Fragen zu beantworten.

Teil 2

Wechselkurs

Der Wechselkurs in Teil 2 ist: **3 Punkte = 1 Euro**

Auktionen

Teil 2 besteht aus nur einer Auktion. In dieser Auktion bieten Sie in Ihrer Gruppe für ein Gut, das einen bestimmten Wert für Sie bzw. Ihre Gruppe hat. Drei Gruppen werden wiederum für die Auktion zufällig verbunden und bilden einen Markt, um für ein Gut zu bieten.

Jedes Gruppenmitglied erhält eine einmalige Anfangsausstattung von 10 Punkten.

Wert des Gutes

Jede Bietergruppe erhält zu Beginn die Information über den Wert des Gutes. Dieser Wert wird für jede Gruppe auf einem Markt unabhängig bestimmt; d.h., es ist sehr wahrscheinlich, dass er für verschiedene Gruppen unterschiedlich ist. Der Wert für jede Gruppe wird zufällig zwischen 0 und 100 Punkten gezogen. Dabei ist jeder Wert zwischen 0 und 100 gleichwahrscheinlich (wobei 0 und 100 auch möglich sind). Jede Bietergruppe kennt ihren eigenen Wert, aber nicht die Werte der beiden anderen Gruppen auf dem Markt.

Entscheidungsfindung

In diesem Teil wird jede Gruppe ein Gebot abgeben. Jede Gruppe entscheidet sich gleichzeitig für ein Gebot; es gibt keine Wiederholung. Die Entscheidung innerhalb der Gruppe erfolgt wieder in zwei Stufen:

In der ersten Stufe werden Sie individuell um einen Vorschlag für das Gruppengebot gefragt. Sie haben 30 Sekunden Zeit, einen Vorschlag in der Mitte des Bildschirms einzugeben und auf "Weiter" zu klicken. Nachdem alle Gruppenmitglieder Ihrer Gruppe den Vorschlag eingegeben haben, startet die zweite Stufe.

In der zweiten Stufe hat die Gruppe 180 Sekunden, um sich auf eine Entscheidung zu verständigen. Eine Entscheidung ist nur dann gültig, wenn von den drei Gruppenmitgliedern jeweils exakt dieselbe Entscheidung eingegeben wird (Entscheidungen Ihrer Gruppenmitglieder sehen Sie im rechten Teil Ihres Bildschirms). Innerhalb der 180 Sekunden können Sie und die anderen Gruppenmitglieder ihre Entscheidungen so oft anpassen, wie Sie wollen. Sie können Ihre Entscheidungen mit den anderen Gruppenmitgliedern auch diskutieren – diese Diskussion erfolgt über ein Chat-Fenster im linken Teil des Bildschirms.

Für den Chat gelten folgende Regeln: (i) deutschsprachige Konversation; (ii) keine

Beleidigungen, Drohungen oder ähnliche Verstöße gegen die Etikette; (iii) keine Informationen, die es erlauben würden Sie zu identifizieren (z.B. Sitzplatznummer, Namen, Geschlecht, Studienrichtung etc.). Wenn Sie gegen diese Regeln verstoßen, werden Sie aus dem Experiment ausgeschlossen.

Der Chat soll Ihnen helfen, sich innerhalb der Gruppe zu koordinieren. Wenn es eine Gruppe nicht schafft, innerhalb von 180 Sekunden eine gemeinsame Entscheidung zu treffen, wird ein Gebot von 0 angenommen.

Gebote müssen immer ganzzahlig sein. Das Höchstgebot beträgt 110 Punkte.

Ende der Auktion

Die Gruppe mit dem höchsten Gebot der drei Gruppen gewinnt die Auktion. Sie bezahlt für das Gut als Preis das zweithöchste Gebot am Markt. Jedes Gruppenmitglied verdient also in dieser Auktion:

Auszahlung jedes Gruppenmitglieds = Wert des Gutes – zweithöchstes Gebot am Markt

Beispiel: Die drei Gruppengebote sind 80, 60, 40; die Gruppe, die 80 Punkte geboten hat, gewinnt die Auktion. Jedes Mitglied dieser Gruppe erhält den Wert des Gutes und bezahlt 60 Punkte.

Wenn alle drei Gruppen 0 geboten haben, wird das Gut nicht verkauft. Wenn zwei oder drei Bietergruppen das gleiche, höchste Gebot (größer 0) abgegeben haben, wird zufällig bestimmt, welche Gruppe das Gut erhält und bezahlt.

Wenn eine Gruppe das Gut in einer Auktion nicht erhält, verdient diese Gruppe in dieser Auktion nichts.

Jedes Gruppenmitglied behält allerdings die Anfangsausstattung von 10 Punkten (mit einer Ausnahme: siehe nächster Absatz).

Achtung!

Sie können auch Verluste in dieser Auktion machen. Wenn Ihre Gruppe einen Preis bezahlt, der über dem Wert des Gutes liegt (= wenn das zweithöchste Gebot höher ist als der Wert des Gutes und Sie die Auktion gewinnen), dann machen Sie Verluste! Verluste werden von Ihrer Anfangsausstattung abgezogen (und wenn diese nicht reicht: von Gewinnen aus anderen Teilen des Experiments).

Es gibt nur eine Auktion in Teil 2.

Zusammenfassung

Nur zwei Dinge ändern sich im Vergleich zu Teil 1:

- Es gibt nur eine Auktion.
- Der Gewinner der Auktion bezahlt das zweithöchste Gebot am Markt für das Gut.

Wenn Sie Fragen haben, dann heben Sie bitte Ihre Hand. Ein Experimentleiter kommt dann zu Ihnen, um Ihre Fragen zu beantworten.

Teil 3

Wechselkurs

Der Wechselkurs in Teil 3 ist: **40 Punkte = 1 Euro**

Jede Gruppe erhält eine Ausstattung von 100 Punkten pro Mitglied. Ihre Gruppe kann jeden Betrag zwischen 0 und 100 Punkten (wobei 0 und 100 auch möglich sind) in eine riskante Anlage investieren und den Rest behalten. Mit R bezeichnen wir die Menge an Punkten, die Ihre Gruppe in die riskante Anlage investiert. Mit 50% Wahrscheinlichkeit wird der Betrag R mit 2,5 multipliziert; mit 50% Wahrscheinlichkeit geht der Betrag R verloren. Jedes Gruppenmitglied erhält also aus dem Investment entweder $2,5 \times R$ oder nichts mit gleicher Wahrscheinlichkeit. Jedes Gruppenmitglied behält den Betrag $100 - R$, der nicht investiert wurde.

Entscheidungsfindung

Die Entscheidung innerhalb der Gruppe erfolgt wieder in zwei Stufen, wobei es sich um dieselbe Gruppe wie in Teil 1 und 2 handelt.

In der ersten Stufe werden Sie individuell um einen Vorschlag für den zu investierenden Betrag R gefragt. Sie haben 30 Sekunden Zeit, einen Vorschlag in der Mitte des Bildschirms einzugeben und auf "Weiter" zu klicken. Wenn Sie keinen Vorschlag machen, bevor die Zeit abgelaufen ist, werden Sie ohne Vorschlag automatisch auf die zweite Stufe weitergeleitet.

In der zweiten Stufe hat die Gruppe 120 Sekunden, um sich auf eine Entscheidung über den zu investierenden Betrag R zu verständigen. Eine Entscheidung ist nur dann gültig, wenn von den drei Gruppenmitgliedern jeweils exakt dieselbe Entscheidung eingegeben wird (Entscheidungen Ihrer Gruppenmitglieder sehen Sie im rechten Teil Ihres Bildschirms). Innerhalb der 120 Sekunden können Sie und die anderen Gruppenmitglieder ihre Entscheidungen so oft anpassen, wie Sie wollen. Sie können Ihre Entscheidungen mit den anderen Gruppenmitgliedern auch diskutieren – diese Diskussion erfolgt über ein Chat-Fenster im linken Teil des Bildschirms.

Für den Chat gelten folgende Regeln: (i) deutschsprachige Konversation; (ii) keine Beleidigungen, Drohungen oder ähnliche Verstöße gegen die Etikette; (iii) keine Informationen, die es erlauben würden Sie zu identifizieren (z.B. Sitzplatznummer, Namen, Geschlecht, Studienrichtung etc.). Wenn Sie gegen diese Regeln verstoßen, werden Sie aus dem Experiment ausgeschlossen.

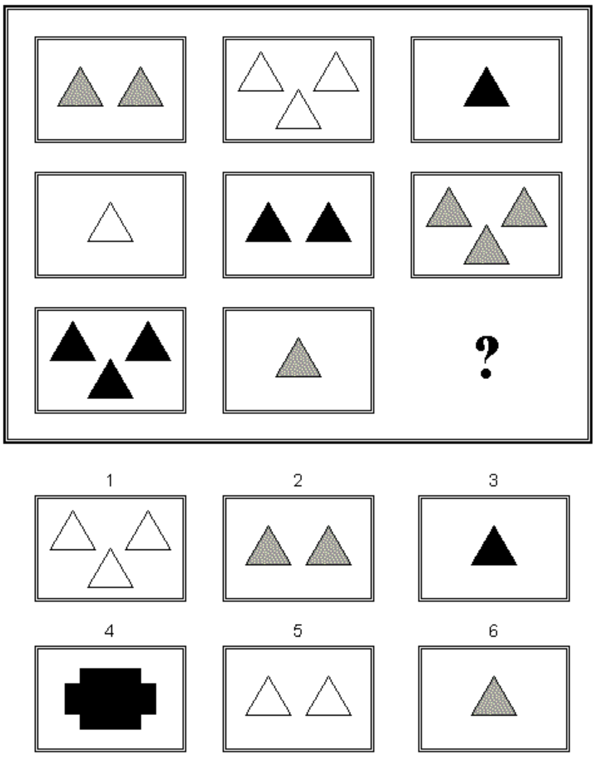
Der Chat soll Ihnen helfen, sich innerhalb der Gruppe zu koordinieren. Wenn es Ihre Gruppe nicht schafft, innerhalb von 120 Sekunden eine gemeinsame Entscheidung zu treffen, verdie-nen Sie in Teil 3 kein Geld.

Der Computer wird nach einer gültigen Entscheidungsfindung die Investmententscheidung simulieren, und Sie erfahren, wie viel Sie verdient haben.

Teil 4

In diesem Teil bitten wir Ihre Gruppe 8 Probleme zu lösen. Alle haben die gleiche Struktur.

Sie sehen einen Kasten mit einer Matrix, d.h. ein rechteckiges Muster verschiedener Symbole. Jede Matrix hat 3 Reihen und 3 Spalten. Das Symbol in der rechten unteren Ecke fehlt. Unter der Matrix sind 6 Symbole zur Auswahl. Genau eines passt in das Muster und sollte das leere Feld ersetzen. Hier ist ein Beispiel.



Die korrekte Lösung ist "Nummer 5". Die Aufgabe Ihrer Gruppe ist es, das korrekte Symbol zu identifizieren. Sobald Sie mit einem Problem fertig sind, erscheint das nächste Problem auf dem Bildschirm. Sie können allerdings nicht zu früheren Problemen zurückkehren, sobald Sie eine Lösung bestätigt haben. Pro korrekte Lösung erhält jedes Gruppenmitglied 50 Cent. Falls die Antwort falsch sein sollte, bekommen Sie nichts.

Die Entscheidung innerhalb der Gruppe erfolgt wieder in zwei Stufen:

In der <u>ersten Stufe</u> werden Sie individuell um einen Vorschlag für die korrekte Symbolnummer gefragt. Sie haben 30 Sekunden Zeit, einen Vorschlag in der Mitte des Bildschirms einzugeben und auf "Weiter" zu klicken. Wenn Sie keinen Vorschlag machen, bevor die Zeit abgelaufen ist, werden Sie ohne Vorschlag automatisch auf die zweite Stufe weitergeleitet.

In der <u>zweiten Stufe</u> hat die Gruppe 60 Sekunden, um sich auf eine Entscheidung über die korrekte Symbolnummer zu verständigen. Eine Entscheidung ist nur dann gültig, wenn von den drei Gruppenmitgliedern jeweils exakt dieselbe Entscheidung eingegeben wird (Entscheidungen Ihrer Gruppenmitglieder sehen Sie im rechten Teil Ihres Bildschirms). Innerhalb der 60 Sekunden können Sie und die anderen Gruppenmitglieder ihre Entscheidungen so oft anpassen, wie Sie wollen. Sie können Ihre Entscheidungen mit den anderen Gruppenmitgliedern auch diskutieren – diese Diskussion erfolgt über ein Chat-Fenster im linken Teil des Bildschirms.

Für den Chat gelten folgende Regeln: (i) deutschsprachige Konversation; (ii) keine Beleidigungen, Drohungen oder ähnliche Verstöße gegen die Etikette; (iii) keine Informationen, die es erlauben würden Sie zu identifizieren (z.B. Sitzplatznummer, Namen, Geschlecht, Studienrichtung etc.). Wenn Sie gegen diese Regeln verstoßen, werden Sie aus dem Experiment ausgeschlossen.

Der Chat soll Ihnen helfen, sich innerhalb der Gruppe zu koordinieren. Wenn es Ihre Gruppe nicht schafft, innerhalb von 60 Sekunden eine gemeinsame Entscheidung zu treffen, verdienen Sie für das betreffende Problem kein Geld.

Am Ende von Teil 4 erfahren Sie, wie viele Probleme Ihre Gruppe korrekt gelöst hat.

Bevor wir Ihnen Ihren Verdienst in bar ausbezahlen, bitten wir Sie noch einen kurzen Frage-bogen auszufüllen. Dann endet das Experiment.

Appendix B Optimal third-degree price discrimination and reciprocity: A theoretical model

B.1 Proofs

Proof of Proposition 2.1.

Consumers maximize their utility functions (1) with respect to the purchased quantity. The maximization problem of consumer 1:

$$\max_{q_1} q_1 - \frac{q_1^2}{2} - p_1 q_1 + \rho_1 \Delta_1(p_1, p_2, \gamma)(p_1 q_1 - p_1(1 - p_1))$$

The kindness term Δ_1 is constant with respect to q_1, so I can substitute for it later. The solution to the maximization problem:

$$q_1 = 1 - p_1 + \rho_1 \Delta_1(p_1, p_2, \gamma) p_1 \tag{6}$$

The maximization problem of consumer 2 is solved analogously.

$$\max_{q_2} a_2 q_2 - \frac{a_2 q_2^2}{2} - p_2 q_2 + \rho_2 \Delta_2(p_1, p_2, \gamma)(p_2 q_2 - p_2(1 - \frac{p_2}{a_2}))$$

$$q_2 = 1 - \frac{p_2}{a_2} + \frac{\rho_2 \Delta_2(p_1, p_2, \gamma) p_2}{a_2} \tag{7}$$

Recall the definition of the kindness term:

$$\begin{aligned}
\Delta_1(p_1, p_2, \gamma) = {} & (1 - p_1) - \frac{(1 - p_1)^2}{2} - p_1(1 - p_1) - \\
& - \left[\gamma \left((1 - p_2) - \frac{(1 - p_2)^2}{2} - p_2(1 - p_2)\right) + \right. \\
& \left. + (1 - \gamma) \left(a_2(1 - \frac{p_2}{a_2}) - \frac{a_2(1 - \frac{p_2}{a_2})^2}{2} - p_2(1 - \frac{p_2}{a_2})\right)\right] \\
\Delta_2(p_1, p_2, \gamma) = {} & a_2(1 - \frac{p_2}{a_2}) - \frac{a_2(1 - \frac{p_2}{a_2})^2}{2} - p_2(1 - \frac{p_2}{a_2}) - \\
& - \left[\gamma \left(a_2(1 - \frac{p_1}{a_2}) - \frac{a_2(1 - \frac{p_1}{a_2})^2}{2} - p_1(1 - \frac{p_1}{a_2})\right) + \right. \\
& \left. + (1 - \gamma) \left((1 - p_1) - \frac{(1 - p_1)^2}{2} - p_1(1 - p_1)\right)\right]
\end{aligned}$$

Substituting for Δ_1 and Δ_2 in (6) and (7) and simplifying yields the formulas in Proposition 2.1.

Proof of Proposition 2.2.

The firm maximizes its profit $\pi(p_1, p_2)$, which is a fourth-degree polynomial in prices:

$$\max_{p_1, p_2} p_1 q_1(p_1, p_2) + p_2 q_2(p_1, p_2),$$

where $q_1(p_1, p_2)$ and $q_2(p_1, p_2)$ are the demand functions defined by Proposition 2.1. I solve this problem numerically for each combination of parameters, with a_2 ranging from 0.05 to 1 with a step of 0.05, γ ranging from 0 to 1 with a step of 0.05, ρ_1 and ρ_2 from 0 to 5 with a step of 0.1. First, I solve first-order conditions:

$$\frac{\partial \pi(p_1, p_2, a_2, \rho_1, \rho_2, \gamma)}{\partial p_1} = 0$$
$$\frac{\partial \pi(p_1, p_2, a_2, \rho_1, \rho_2, \gamma)}{\partial p_2} = 0.$$

In each of the 1092420 cells, I obtain several solutions in real numbers. Second, for each solution I check second-order conditions:

$$\frac{\partial^2 \pi(p_1, p_2, a_2, \rho_1, \rho_2, \gamma)}{\partial p_1^2} < 0$$
$$\det H\pi(p_1, p_2, a_2, \rho_1, \rho_2, \gamma) > 0,$$

where H is the Hessian of the profit function. If a solution satisfies second-order conditions, I store it in a vector of potential candidates for optimal prices. Finally, I check the length of this vector in each cell. In every cell there exists only one price pair (length of 2), being the unique solution to the optimization problem. Mathematica code of this proof can be found in the input cells [3] and [4] in Appendix B.2.

Proof of Propositions 2.3, 2.4 and 2.5.

To prove comparative statics, I evaluate the partial derivatives of the interpolated equilibrium prices with respect to the four parameters. Among the values of the partial derivatives on the grid, I look for positive or, respectively, negative values that correspond to the specified range of parameters. For instance, Proposition 2.3 claims that the derivative of p_1 with respect to ρ_2 is non-negative for $a_2 \geq 0.55$. Thereby, I restrict the grid to the parameters that satisfy the latter condition, and search for any negative values of $\partial p_1 / \partial \rho_2$. When I find no negative values on the grid, I claim that the derivative is non-negative for $a_2 \geq 0.55$. Refer to Mathematica input cells [6]–[12] for the proof of Proposition 2.3; input cells [13], [14] for the proof of Proposition 2.4; and input cells [15], [16] for the proof of Proposition 2.5 in Appendix B.2.

There is one case in Proposition 2.3 where I find several positive values of the partial derivative of p_2 with respect to ρ_2 on the respective parameter grid. Namely, this happened for some values of ρ_2 given $a_2 = 0.5$, $\rho_1 = 0$ and $\gamma = 1$ (see input cells [8], [9] and output cell [9]). Inspecting this case closer and looking at the optimal price on the second market, I find that $p_2 = 0.25$ if $a_2 = 0.5$, $\rho_1 = 0$ and $\gamma = 1$ irrespective of the value of ρ_2 (refer to input cell [10] and output cell [10]). Thus, the derivative of p_2 with respect to ρ_2 should be equal to zero in all of these cases. Now let us look at the estimates of the partial derivative in output cell [11]: some of them are precisely equal to 0, and some of them are very close to 0 within the neighborhood of 10^{-15}. I can conclude that the positive values I observe here are merely due to a negligible interpolation error.

B.2 Mathematica code

```
In[1]:= (***MODEL SETUP***)
ClearAll[a2,ρ1,ρ2,γ];
PayoffH[q1_,p1_]=q1- q1^2/2-p1 q1;
PayoffL[q2_,p2_]=a2 q2-a2 q2^2/2-p2 q2;
PayoffF[q1_,q2_,p1_,p2_]=p1 q1+p2 q2;
(*STANDARD PREFERENCES// NO INFO treatment*)
(*consumer demands*)
solqH=Solve[D[PayoffH[q1,p1],q1]==0,q1];
demandHNoInfo[p1_]=q1/.solqH[[1]]//Simplify;
solqL=Solve[D[PayoffL[q2,p2],q2]==0,q2];
demandLNoInfo[p2_]=q2/.solqL[[1]]//Simplify;
(*price discrimination*)
solp=Solve[D[PayoffF[demandHNoInfo[p1],demandLNoInfo[p2],p1,p2
],p1]==0 && D[PayoffF[demandHNoInfo[p1],demandLNoInfo[p2],p1,p2
],p2]==0,{p1,p2}];
pHNoInfo[a2_]=p1/.solp[[1]];
pLNoInfo[a2_]=p2/.solp[[1]];
(*uniform pricing*)
solpuni=Solve[D[PayoffF[demandHNoInfo[p],demandLNoInfo[p],p,p],
p]==0,p];
pUni=p/.solpuni[[1]];

In[2]:= (*RECIPROCAL PREFERENCES// FULL INFO treatment*)
recipH[q1_,p1_]=p1 q1-p1 demandHNoInfo[p1];
recipL[q2_,p2_]= p2 q2-p2 demandLNoInfo[p2];
kindnessH[p1_,p2_]=PayoffH[demandHNoInfo[p1],p1]-(γ PayoffH[
demandHNoInfo[p2],p2]+(1-γ)PayoffL[demandLNoInfo[p2],p2]);
kindnessL[p1_,p2_]=PayoffL[demandLNoInfo[p2],p2]-(γ PayoffL[
demandLNoInfo[p1],p1]+(1-γ)PayoffH[demandHNoInfo[p1],p1]);
uHFullInfo[q1_,p1_,p2_]=PayoffH[q1,p1]+ρ1 kindnessH[p1,p2]
recipH[q1,p1];
uLFullInfo[q2_,p1_,p2_]=PayoffL[q2,p2]+ρ2 kindnessL[p1,p2]
recipL[q2,p2];
```

```
(*consumer demands*)
solqH1fi=Solve[D[uHFullInfo[q1,p1,p2],q1]==0,q1];
demandHFullInfo[p1_,p2_,γ_]=q1/.solqH1fi[[1]];
solqL1fi=Solve[D[uLFullInfo[q2,p1,p2],q2]==0,q2];
demandLFullInfo[p1_,p2_,γ_]=q2/.solqL1fi[[1]];
(*firm's profit*)
ProfitFI[p1_,p2_,a2_,ρ1_,ρ2_,γ_]=Collect[PayoffF[
demandHFullInfo[p1,p2,γ],demandLFullInfo[p1,p2,γ],p1,p2],{p1,
p2}];
```

In[3]:=
```
(*Solve first-order conditions for each combination of
parameters on a grid*)
tab=Flatten[Table[{{a2,ρ1,ρ2,γ}, NSolve[D[ProfitFI[p1,p2,a2,
ρ1,ρ2,γ],p1]==0 && D[ProfitFI[p1,p2,a2,ρ1,ρ2,γ],p2]==0,{p1,p2
},Reals]},{a2,0.2,1,.2},{ρ1,0,5,1},{ρ2,0,5,1},{γ,0,1,.2}],3];
```

In[4]:=
```
(***Check second-order conditions***)
(***second-order derivative of the objective function***)
D2p1[p1_,p2_,{a2_,ρ1_,ρ2_,γ_}]=D[ProfitFI[p1,p2,a2,ρ1,ρ2,γ],
{p1,2}];
(***Hessian***)
Hes[p1_,p2_,{a2_,ρ1_,ρ2_,γ_}]=D[ProfitFI[p1,p2,a2,ρ1,ρ2,γ],
{{p1,p2},2}];
optp=NULL;
notunique=0;
notexist=0;
For[i=1,
i≤Length[tab[[All,1]]],
i++,
vec0={};
For[j=1,
j<=Length[tab[[i,2]]],
j++,
```

```
vec=If[D2p1[tab[[i,2]][[All,1,2]][[j]],tab[[i,2]][[All,2,2]][[j
]],tab[[i,1]]]<0&&Det[Hes[tab[[i,2]][[All,1,2]][[j]],tab[[i
,2]][[All,2,2]][[j]],tab[[i,1]]]]>0,
{tab[[i,2]][[All,1,2]][[j]],tab[[i,2]][[All,2,2]][[j]]},{}];
vec0=Join[vec0,vec]
];
notunique=If[Length[vec0]>2,notunique+1,notunique];
notexist=If[Length[vec0]==0,notexist+1,notexist];
optp=ArrayFlatten[{{optp},{vec0}}]
]
optp=Delete[optp,1];
(***Existence & uniqueness***)
Print["Number of cells with non-unique solutions=",notunique]
Print["Number of cells with non-existent solutions=",notexist]
```

```
Number of cells with non-unique solutions=0
Number of cells with non-existent solutions=0
```

In[5]:=

```
(***interpolate price functions***)
p1opt={};
For[i=1,
i≤ Length[optp],
i++,
p1optadd=If[Length[Flatten[optp[[i]]]]==2,Flatten[optp[[i
]]][[1]],{}];
p1opt=ArrayFlatten[{{p1opt},{p1optadd}}]
]
p1opt=Delete[p1opt,1];
p1tab=ArrayFlatten[Transpose[{{Transpose[{tab[[All,1]]}]},
{p1opt}}]];
p1FullInfoNum=Interpolation[p1tab,ExtrapolationHandler→{(
Indeterminate&),WarningMessage→False}];
p2opt={};
For[i=1,
i≤ Length[optp],
```

```
i++,
p2optadd=If[Length[Flatten[optp[[i]]]]==2,Flatten[optp[[i
]]][[2]],{}];
p2opt=ArrayFlatten[{{p2opt},{p2optadd}}]
]
p2opt=Delete[p2opt,1];
p2tab=ArrayFlatten[Transpose[{{Transpose[{tab[[All,1]]}]},
{p2opt}}]];
p2FullInfoNum=Interpolation[p2tab,ExtrapolationHandler→{(
Indeterminate&),WarningMessage→False}];
```

In[6]:=
```
(*Proof: p1 decreases in rho1*)
(*condition: gamma>0 and rho1-rho2<=3.5*)
Dp1rho1[a2_,ρ1_,ρ2_,γ_]=D[p1FullInfoNum[a2,ρ1,ρ2,γ],ρ1];
tabDp1rho1=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp1rho1[a2,ρ1,ρ2,γ]},
{a2,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp1rho1,MemberQ[#,_?Positive]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: γ>0 & ρ1-ρ2≤3.5"]
Print["Number of points where ∂p1/∂ρ1 is positive=",Length[Select[
val,#[[4]]>0&&#[[2]]-#[[3]]≤3.5&]]]
```

Condition: $\gamma>0$ & $\rho_1-\rho_2\leq3.5$

Number of points where $\frac{\partial p_1}{\partial \rho_1}$ is positive=0

In[7]:=
```
(*Proof: p1 increases in rho2*)
(*condition: a2>=0.55*)
Dp1rho2[a2_,ρ1_,ρ2_,γ_]=D[p1FullInfoNum[a2,ρ1,ρ2,γ],ρ2];
tabDp1rho2=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp1rho2[a2,ρ1,ρ2,γ]},
{a2,0.1,1,.1},{ρ1,0,10,.5},{ρ2,0,10,.5},{γ,0,1,.1}],3];
val=Select[tabDp1rho2,MemberQ[#,_?Negative]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: a2≥0.55"]
Print["Number of points where ∂p1/∂ρ2 is negative=",Length[Select[
val,#[[1]]≥0.55&]]]
```

```
Condition: a2≥0.55
Number of points where ∂p1/∂ρ2 is negative=0
```

In[8]:=
```
(*Proof: p2 decreases in rho2*)
(*condition: rho2-rho1<=3.5*)
Dp2rho2[a2_,ρ1_,ρ2_,γ_]=D[p2FullInfoNum[a2,ρ1,ρ2,γ],ρ2];
tabDp2rho2=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp2rho2[a2,ρ1,ρ2,γ]},
{a2,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp2rho2,MemberQ[#,_?Positive]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: ρ2-ρ1≤3.5"]
Print["Number of points where ∂p2/∂ρ2 is positive=",Length[Select[
val,#[[3]]-#[[2]]≤3.5&]]]
```

```
Condition: ρ2-ρ1≤3.5
Number of points where ∂p2/∂ρ2 is positive=12
```

In[9]:=
```
Select[val,#[[3]]-#[[2]]≤3.5&]
```

Out[9]=
```
{{0.5,0.,0.,1.},{0.5,0.,0.3,1.},{0.5,0.,0.5,1.},{0.5,0.,0.6,1.
},{0.5,0.,0.9,1.},{0.5,0.,1.,1.},{0.5,0.,1.1,1.},{0.5,0.,1.3,1.
},{0.5,0.,1.5,1.},{0.5,0.,1.9,1.},{0.5,0.,2.8,1.},
{0.5,0.,3.3,1.}}
```

In[10]:=
```
Table[p2FullInfoNum[0.5,0,ρ2,1],{ρ2,0,5,.1}]
Table[Dp2rho2[0.5,0,ρ2,1],{ρ2,0,5,.1}]
```

Out[10]=
```
{0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25, 0.25,
0.25}
```

```
Out[11]= {1.6653345369377348`*10^-15, -2.7755575615628914`*10^-16,
-5.551115123125783`*10^-16, 9.248244531301353`*10^-17, 0.`,
1.850733108432756`*10^-16, 2.7750154604766486`*10^-16,
-7.401306130472296`*10^-16, -1.8501910073465133`*10^-16,
5.551115123125783`*10^-16, 9.25366554216378`*10^-17,
2.7755575615628914`*10^-16, -3.701466216865512`*10^-16,
3.7009241157792694`*10^-16, -5.551115123125783`*10^-16,
9.25366554216378`*10^-17, 0.`, -2.7755575615628914`*10^-16,
-4.163336342344337`*10^-16, 8.326672684688674`*10^-16,
-1.3877787807814457`*10^-16, 0.`, 0.`, 0.`, 0.`,
-9.25366554216378`*10^-17, -3.237969788127959`*10^-16, 0.`,
5.088702896560715`*10^-16, -9.25366554216378`*10^-17, 0.`,
-1.8501910073465133`*10^-16, -2.7755575615628914`*10^-16,
5.551115123125783`*10^-16, -9.25366554216378`*10^-17, 0.`,
-2.7755575615628914`*10^-16, -4.163336342344337`*10^-16,
8.326672684688674`*10^-16, -1.3877787807814457`*10^-16, 0.`, 0.`,
0.`, 0.`, 0.`, 0.`, -9.25366554216378`*10^-17,
-1.388049831324567`*10^-16, 2.77528651101977`*10^-16,
-4.629543276513104`*10^-17, 9.25366554216378`*10^-17}

In[12]:= (*Proof: p2 increases in rho1*)
(*condition: always*)
Dp2rho1[a2_,ρ1_,ρ2_,γ_]=D[p2FullInfoNum[a2,ρ1,ρ2,γ],ρ1];
tabDp2rho1=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp2rho1[a2,ρ1,ρ2,γ]},
{a2,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp2rho1,MemberQ[#,_?NonPositive]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: always"]
Print["Number of points where ∂p2/∂ρ1 is nonpositive=",Length[val
]]

Condition: always
Number of points where ∂p2/∂ρ1 is nonpositive=0

In[13]:= (*Proof: p1 increases in a2*)
```

```
(*condition: gamma>=0.4 and rho1>=rho2 and a2>=0.1*)
Dp1a2[a2_,ρ1_,ρ2_,γ_]=D[p1FullInfoNum[a2,ρ1,ρ2,γ],a2];
tabDp1a2=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp1a2[a2,ρ1,ρ2,γ]},{a2
,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp1a2,MemberQ[#,_?Negative]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: γ ≥0.4 and ρ1≥ρ2 and a2≥0.1"]
Print["Number of points where ∂p1/∂a2 is negative=",Length[Select[
val,#[[4]]>=0.4&&#[[2]]>=#[[3]]&&#[[1]]>=0.1&]]]
```

Condition: $\gamma \geq 0.4$ and $\rho_1 \geq \rho_2$ and $a_2 \geq 0.1$

Number of points where $\frac{\partial p_1}{\partial a_2}$ is negative=0

In[14]:=

```
(*Proof: p2 increases in a2*)
(*condition: always*)
Dp2a2[a2_,ρ1_,ρ2_,γ_]=D[p2FullInfoNum[a2,ρ1,ρ2,γ],a2];
tabDp2a2=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp2a2[a2,ρ1,ρ2,γ]},{a2
,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp2a2,MemberQ[#,_?NonPositive]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: always"]
Print["Number of points where ∂p2/∂a2 is nonpositive=",Length[val
]]
```

Condition: always

Number of points where $\frac{\partial p_2}{\partial a_2}$ is nonpositive=0

In[15]:=

```
(*Proof: p1 decreases in gamma*)
(*condition: a2<1 and rho1>0*)
Dp1g[a2_,ρ1_,ρ2_,γ_]=D[p1FullInfoNum[a2,ρ1,ρ2,γ],γ];
tabDp1g=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp1g[a2,ρ1,ρ2,γ]},{a2
,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp1g,MemberQ[#,_?Positive]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: a2<1 and ρ1>0"]
```

```
Print["Number of points where ∂p1/∂γ is positive=",Length[Select[
val,#[[1]]<1&&#[[2]]>0&]]]
```

Condition: a_2<1 and ρ_1>0

Number of points where $\frac{\partial p_1}{\partial \gamma}$ is positive=0

In[16]:=
```
(*Proof: p2 increases in gamma*)
(*condition: 0.55<=a2<1 and gamma>=0.2 and rho2>0 and rho1-rho2
<=3*)
Dp2g[a2_,ρ1_,ρ2_,γ_]=D[p2FullInfoNum[a2,ρ1,ρ2,γ],γ];
tabDp2g=Flatten[Table[{{a2,ρ1,ρ2,γ}, Dp2g[a2,ρ1,ρ2,γ]},{a2
,0.1,1,.1},{ρ1,0,5,.5},{ρ2,0,5,.5},{γ,0,1,.1}],3];
val=Select[tabDp2g,MemberQ[#,_?Negative]&];
val=Flatten[Take[val,All,1],1];
Print["Condition: 0.55≤a2<1 and γ ≥0.2 and ρ2>0 and ρ1-ρ2≤3"]
Print["Number of points where ∂p2/∂γ is negative=",Length[Select
[val,0.55<=#[[1]]<1&&#[[4]]≥0.2&&#[[3]]>0&&#[[2]]-#
[[3]]<=3&]]]
```

Condition: 0.55≤a_2<1 and γ ≥0.2 and ρ_2>0 and ρ_1-ρ_2≤3

Number of points where $\frac{\partial p_2}{\partial \gamma}$ is negative=0

Appendix C Optimal third-degree price discrimination and reciprocity: Experimental evidence

C.1 Proofs of hypotheses

Theoretical foundations behind Hypotheses 3.1, 3.2, 3.3 and 3.4.

To obtain consumer demands given the two pricing schemes, we evaluate the demand functions presented in Proposition 2.1 at (p_1^{PD}, p_2^{PD}) and (p_1^{UNI}, p_2^{UNI}). The resulting functions of parameters a_2, ρ_1, ρ_2 and γ correspond to $q_i^{PD}(Full\ Info)$ and $q_i^{PD}(Full\ Info)$, respectively. If we evaluate these functions at $\gamma = 1$, we will arrive at $q_i^{PD}(Part\ Info)$ and $q_i^{PD}(Part\ Info)$. Evaluating them at $\gamma = 1,\ \rho_1 = \rho_2 = 0$ gives us the demands in *No Info* treatment. To estimate the treatment effects, we evaluate and simplify the respective differences. First, let us provide the basis for Hypothesis 3.1.

$$q_1^{PD}(Full\ Info) - q_1^{PD}(Part\ Info) = \frac{1}{16}(1-a_2)(4-a_2)(1-\gamma)\rho_1 \tag{8}$$

$$q_1^{PD}(No\ Info) - q_1^{PD}(Part\ Info) = \frac{1}{16}(1-a_2)(3-a_2)\rho_1 \tag{9}$$

$$q_1^{PD}(No\ Info) - q_1^{PD}(Full\ Info) = \frac{1}{16}(1-a_2)(-1+4\gamma-a_2\gamma)\rho_1 \tag{10}$$

Recall that $a_2 < 1$ by construction. Difference (8) is always nonnegative, and positive if $\gamma < 1$ and $\rho_1 > 0$. Difference (9) is always nonnegative, and positive if $\rho_1 > 0$. Difference (10) is nonnegative if $\gamma \geq \frac{1}{4-a_2}$, and positive if the latter condition is non-binding and $\rho_1 > 0$.

$$q_2^{PD}(Full\ Info) - q_2^{PD}(Part\ Info) = \frac{1}{16a_2}(1-a_2)(1-4a_2)(1-\gamma)\rho_2 \tag{11}$$

$$q_2^{PD}(No\ Info) - q_2^{PD}(Part\ Info) = \frac{1}{16a_2}(1-a_2)(1-3a_2)\rho_2 \tag{12}$$

$$q_2^{PD}(No\ Info) - q_2^{PD}(Full\ Info) = \frac{1}{16a_2}(1-a_2)(a_2+\gamma-4a_2\gamma)\rho_2 \tag{13}$$

Hypothesis 3.2 is derived analogously. Difference (11) is always nonpositive since we restricted $a_2 \geq 0.5$ in the experiment, and negative if $\gamma < 1$ and $\rho_2 > 0$. Difference

(12) is nonpositive as well, and negative when $\rho_2 > 0$. Difference (13) is nonpositive if $\gamma \geq \frac{a_2}{4a_2-1}$, and negative when the latter condition is non-binding and $\rho_2 > 0$.

Finally, it is obvious that the demands of both consumers should not differ across *No Info* and *Part Info* treatments when they are charged the same price. In both treatments the model predicts rational demands since the kindness terms would amount to zero in *Part Info* treatment. The rest of Hypotheses 3.3 and 3.4 compares the demands under *Full Info* and *No Info* treatments.

$$q_1^{UNI}(Full\ Info) - q_1^{UNI}(No\ Info) = \frac{a_2(1-a_2)(1+a_2+a_2^2)(1-\gamma)\rho_1}{2(1+a_2)^3} \tag{14}$$

$$q_2^{UNI}(Full\ Info) - q_2^{UNI}(No\ Info) = -\frac{(1-a_2)(1+a_2+a_2^2)(1-\gamma)\rho_2}{2(1+a_2)^3} \tag{15}$$

Difference (14) is always nonnegative, and positive if $\gamma < 1$ and $\rho_1 > 0$. Difference (15) is always nonpositive, and negative if $\gamma < 1$ and $\rho_2 > 0$. The comparison of consumer demands across the three treatments and the two pricing schemes is complete.

Theoretical foundations behind Hypothesis 3.5.

To assess the attractiveness of price discrimination for the firm in each treatment, we need to compare the profits obtained under the two pricing schemes. We plug in p_1^{PD}, p_2^{PD}, p^{UNI} and the treatment-specific q_i^{PD}, q_i^{UNI} into the firm's profit function. Consequently, we evaluate the relative benefits of price discrimination $\pi^{PD} - \pi^{UNI}$ and compare them across treatments.

First, note that $\pi^{PD}(Full\ Info) - \pi^{UNI}(Full\ Info)$ is linear and (weakly) decreasing in γ:

$$\frac{\partial}{\partial\gamma}\left(\pi^{PD}(Full\ Info) - \pi^{UNI}(Full\ Info)\right) =$$
$$= -\frac{(1-a_2)^2\left((4+19a_2+23a_2^2+17a_2^3+a_2^4)\rho_1 + (1+17a_2+23a_2^2+19a_2^3+4a_2^4)\rho_2\right)}{32(1+a_2)^4} \leq 0.$$

We assume that in *Part Info* treatment $\gamma = 1$. Since relative attractiveness of price discrimination in *Full Info* treatment decreases in γ, the attractiveness of price discrimination in *Part Info* treatment will be lower or equal to that in *Full Info* treatment, all other parameters being held constant.

Second, consider now the relative gain from price discrimination in *Part Info* treatment. It is linear in both reciprocity parameters, decreasing in ρ_1 and increasing in ρ_2 given $a_2 \geq 0.5$:

$$\frac{\partial}{\partial \rho_1}\left(\pi^{PD}(Part\ Info) - \pi^{UNI}(Part\ Info)\right) = -\frac{1}{32}(3 - a_2)(1 - a_2) < 0; \qquad (16)$$

$$\frac{\partial}{\partial \rho_2}\left(\pi^{PD}(Part\ Info) - \pi^{UNI}(Part\ Info)\right) = \frac{1}{32}(3a_2 - 1)(1 - a_2) > 0. \qquad (17)$$

The absolute value of (16) is higher than (17), meaning that the relative profitability of price discrimination decreases in ρ_1 faster than it increases in ρ_2. If we assume that $\rho_1 > \rho_2$, it guarantees that price discrimination is strictly more attractive in *No Info* treatment than in *Part Info* treatment since we assume $\rho_1 = \rho_2 = 0$ in *No Info* treatment.

The only comparison that remains open is *Full Info* versus *No Info* treatment. The latter treatment shuts down the effect of two reciprocity parameters and γ simultaneously. Relative attractiveness of price discrimination is (weakly) decreasing in γ and ρ_1 and increasing in ρ_2. The sum of the three effects would depend on certain parameter values, making the assessment of the relative attractiveness of price discrimination more difficult.

Recall that the predicted share of price discriminating firms in *No Info* treatment amounts to 1 since the behavior of consumers is expected to be selfish-rational. Thereby, it puts an upper bound on the rates of price discrimination in *Full Info* treatment: they can be either also equal to 1 or lower, depending on the parameters. The share of price discriminating firms in *Part Info* will thus be the lowest in our experiment.

C.2 Instructions

Paper-based instructions for all subjects in *Full Info* treatment are provided below (in German).

Herzlich willkommen beim Experiment und vielen Dank für Ihre Teilnahme!

Bitte sprechen Sie ab jetzt nicht mehr mit den anderen Teilnehmern.

Der sprachlichen Einfachheit halber werden im Folgenden männliche Bezeichnungen verwendet; diese schließen sowohl männliche als auch weibliche Teilnehmer mit ein.

Allgemeines zum Ablauf

Dieses Experiment dient der Untersuchung von ökonomischem Entscheidungsverhalten. Sie können dabei Geld verdienen, das Ihnen im Anschluss an das Experiment privat und in bar ausbezahlt wird. Während des Experiments werden Sie darum gebeten, Entscheidungen zu treffen. Zum Teil werden Sie dabei mit anderen Teilnehmern interagieren. Das heißt, dass sowohl Ihre eigenen Entscheidungen als auch die Entscheidungen der anderen Teilnehmer Ihre Auszahlung beeinflussen können. Diese ergibt sich entsprechend bestimmter Regeln, die im Folgenden erklärt werden.

Das gesamte Experiment dauert etwa 90 Minuten. Während Sie Ihre Entscheidungen treffen, wird am rechten oberen Rand des Bildschirms eine Uhr herunterlaufen. Diese bietet Ihnen eine Orientierung, wie viel Zeit Sie für Ihre Entscheidungen benötigen.

Bezahlung

Während des Experiments berechnen sich Verdienste nicht in Euro, sondern in Punkten. Am Ende des Experiments werden die Punkte, die Sie verdient haben, in Euro umgerechnet und Ihnen privat und in bar ausbezahlt. Der Wechselkurs für die Umrechnung beträgt: **150 Punkte = 1 Euro**

Zusätzlich zu dem Einkommen, das Sie während des Experimentes verdienen können, erhalten Sie 6 ď für Ihr pünktliches Erscheinen.

Anonymität

Die Daten aus dem Experiment werden ausschließlich anonym ausgewertet. Am Ende des Experiments müssen Sie eine Quittung über den Erhalt des Verdienstes unterschreiben. Diese dient nur der Abrechnung und wird nicht dazu verwendet, Ihre persönlichen Daten mit Ihren Entscheidungen zu verknüpfen. Ihr Name kann zu keinem Zeitpunkt mit Ihrem Verhalten im Experiment kombiniert werden.

Hilfsmittel

An Ihrem Platz finden Sie einen Kugelschreiber. Bitte lassen Sie diesen nach dem Experiment auf dem Tisch liegen.

Sollten Sie nach den Instruktionen oder während des Experiments noch Fragen haben, drücken Sie bitte den roten Knopf auf Ihrer Tastatur (F11). Der Experimentleiter wird dann zu Ihnen kommen und Ihre Fragen privat beantworten. Wenn Sie keine Hilfe mehr benötigen, drücken Sie einfach erneut auf den roten Knopf.

Ablauf des Experiments

Märkte

Zu Beginn des Experiments wird Ihnen zufällig eine Rolle zugewiesen: Verkäufer oder Käufer. Diese Einteilung bleibt über das gesamte Experiment hinweg bestehen. Zu Beginn jeder Runde werden alle Teilnehmer zufällig in Gruppen eingeteilt, die jeweils aus einem Verkäufer und zwei Käufern bestehen. Wir nennen jede Dreiergruppe einen Markt.

Ablauf einer Runde

Insgesamt besteht das Experiment aus 20 unabhängigen Runden und einer Proberunde. Der Verkäufer kann ein Gut an beide Käufer auf seinem Markt verkaufen. In jeder Runde entscheiden die Käufer, welche Menge des Gutes sie kaufen möchten. Der Verkäufer entscheidet, zu welchen Preisen er das Gut verkauft. Die Entscheidungsmöglichkeiten von Verkäufern und Käufern sind im Folgenden ausführlich erklärt.

Entscheidung des Verkäufers

Zu Beginn jeder Runde wählt der Verkäufer zwei Preise für die zwei Käufer auf seinem Markt. Dabei kann er immer aus zwei Preispaaren auswählen.

Der Verkäufer hat 120 Sekunden Zeit, sich für eines der zwei gegebenen Preispaare zu entscheiden.

Entscheidung der Käufer

Jeder Käufer entscheidet sich, welche Menge des Gutes er kaufen will.

Der Käufer trifft diese Entscheidung für beide möglichen Preise, die der Verkäufer von ihm verlangen kann (es gibt also zwei Entscheidungen pro Runde). Erst nach der Entscheidung erfährt jeder Käufer, welchen Preis der Verkäufer tatsächlich festgesetzt hat. Nur die Menge, die zum tatsächlich festgesetzten Preis gewählt wurde, wird für die Berechnung der Auszahlungen berücksichtigt. Allerdings weiß der Käufer zum Zeitpunkt der Entscheidung natürlich noch nicht, welcher der Preise tatsächlich gewählt wurde und welche seiner Mengenentscheidungen relevant ist. Daher sollte der Käufer in jeder Situation alle Mengenentscheidungen sorgfältig bedenken, da jede einzelne potenziell relevant für seine Auszahlung werden kann.

Die Käufer haben ebenfalls 120 Sekunden Zeit, um sich für Mengen zu entscheiden.

Verdiensttabellen

Käufer und Verkäufer können ihre möglichen Verdienste in Tabellen auf dem Bildschirm ablesen, wenn sie ihre Entscheidungen treffen.

Eine Beispieltabelle für den Käufer sieht aus wie folgt:

Menge	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
Ihr Verdienst, wenn: Ihr Preis = **5**	0	82	153	212	260	297	323	337	340	332	313	282	240	187	123	47	0
Ihr Verdienst, wenn: Ihr Preis = **6**	0	77	143	197	240	272	293	302	300	287	263	227	180	122	53	0	0

In dieser Tabelle kann der Käufer die möglichen Verdienste abhängig von der von ihm gewählten Menge für die zwei möglichen Preise ablesen. Je höher der Preis ist, den der Käufer erhält, desto niedriger sind die möglichen Verdienste.

Betrachten Sie die Beispieltabelle oben. Der Käufer kann hier zwischen den Mengen 0, 5, 10, ..., 80 wählen. Diese stehen in der ersten Zeile. Die folgenden zwei Zeilen zeigen die Verdienste des Käufers für die zwei möglichen Preise in dieser Runde: 5 und 6.

Außerdem erfährt der Käufer die Preise, die der Verkäufer für den anderen Käufer auf seinem Markt festlegen kann, und erhält eine weitere Tabelle, die den Verdienst des anderen Käufers für diese Preise zeigt.

Der Verkäufer erhält alle Informationen aus den Tabellen, die die Käufer auf seinem Markt erhalten. Eine Beispieltabelle für den Verkäufer sieht dabei wie folgt aus:

Menge, die Käufer 1 wählt	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80
Verdienst von Käufer 1 (sein Preis = **5**)	0	82	153	212	260	297	323	337	340	332	313	282	240	187	123	47	0
Ihr Verdienst aus dem Verkauf an Käufer 1	0	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400

Die Tabelle zeigt die möglichen Verdienste abhängig von der Entscheidung von Käufer 1, wenn er einen Preis von 5 erhält.

Die erste Zeile zeigt die möglichen Mengen, die Käufer 1 kaufen kann. Die zugehörigen Verdienste von Käufer 1 stehen in Zeile 2. In der dritten Zeile steht der Verdienst des Verkäufers aus dem Verkauf des Gutes an Käufer 1.

Der Verdienst des Verkäufers aus dem Verkauf an Käufer 1 entspricht dem Preis für Käufer 1 multipliziert mit der Menge, die Käufer 1 kauft. Betrachten Sie erneut das Beispiel oben. Bei einem Preis von 5 für Käufer 1 und einer gekauften Menge von 30 für diesen Preis, ergibt sich der Verdienst aus dem Verkauf des Gutes an Käufer 1 als 5*30 = 150. Diese Zahl findet der Verkäufer in der untersten Zeile.

Der gesamte Verdienst des Verkäufers in einer Runde entspricht der Summe, die er aus dem Verkauf des Gutes an die zwei Käufer in seinem Markt verdient:

$$\text{Verdienst} = \text{Preis für Käufer 1} * \text{Menge von Käufer 1} +$$
$$+ \text{Preis für Käufer 2} * \text{Menge von Käufer 2}$$

Beachten Sie, dass der Verdienst des Verkäufers von den gekauften Mengen abhängt. Je mehr die Käufer zu einem bestimmten Preis kaufen, desto mehr verdient der Verkäufer.

Feedback in einer Runde

Nachdem der Verkäufer und die zwei Käufer auf einem Markt ihre Entscheidungen getroffen haben, erfahren sie ihre jeweiligen Verdienste in dieser Runde.

Feedback für den Verkäufer

Am Ende der Runde wird dem Verkäufer angezeigt, welche Mengen er jeweils tatsächlich verkauft hat. Darüber hinaus erfährt er sowohl seinen Verdienst aus dieser Runde als auch die Verdienste der beiden Käufer.

Feedback für den Käufer

Am Ende der Runde erfährt der Käufer, welchen Preis der Verkäufer für ihn tatsächlich festgelegt hat und welche Menge er dementsprechend gekauft hat. Darüber hinaus erfährt er auch seinen Verdienst aus dieser Runde sowie den Preis, den der Verkäufer für den anderen Käufer auf dem Markt festgelegt hat.

Weitere Runden

Ist eine Runde beendet, so werden die Märkte zufällig neu zusammengestellt und es startet eine neue Runde mit exakt demselben Ablauf. Die Rollen (Käufer bzw. Verkäufer) bleiben dabei unverändert. Die Preisoptionen und Tabellen ändern sich dagegen von Runde zu Runde. Nehmen Sie sich daher in jeder Runde die Zeit, die Preisoptionen und Tabellen zu studieren.

In diesem Experiment gibt es insgesamt eine Proberunde und 20 Runden. Die Proberunde ist weiter unten erklärt.

Auszahlungen

Am Ende des Experiments wird eine der 20 Runden zufällig als auszahlungsrelevant ausgewählt (die Proberunde kann nicht ausgewählt werden). Ihre Auszahlung ergibt sich aus den Punkten, die Sie in der ausgewählten Runde verdient haben. Die Verdienste aus allen anderen Runden sind nicht auszahlungsrelevant.

Proberunde

Die Proberunde gibt Ihnen Gelegenheit, sich mit dem Ablauf der folgenden Runden und mit dem Aufbau der Bildschirme vertraut zu machen.

Zunächst erfahren Sie Ihre Rolle (Käufer bzw. Verkäufer). Wie oben bereits erklärt, behalten Sie diese Rolle für den Rest des Experimentes.

Im Anschluss sehen Sie einen Beispielbildschirm. Anhand dieses Beispielbildschirms wird Ihnen noch einmal genau erklärt, welche Informationen Sie in jeder Runde bekommen, wo Sie diese auf dem Bildschirm finden, und wie Sie Ihre Entscheidungen treffen können.

Danach haben Sie die Gelegenheit, Ihre Entscheidungen einmal probeweise zu treffen. In dieser Proberunde erfahren Sie nicht, welche Entscheidungen die anderen Teilnehmer auf Ihrem Markt getroffen haben, sondern erhalten stattdessen ein zufälliges, computergeneriertes Feedback. Diese Runde ist nicht auszahlungsrelevant, sondern dient nur dazu, Sie mit der Entscheidungssituation vertraut zu machen. Sie haben dafür 120 Sekunden Zeit.

Wenn die Proberunde zu Ende ist, startet das aus 20 Runden bestehende Experiment. Danach bitten wir Sie, noch einen kurzen Fragebogen auszufüllen.

Sollten Sie während der Proberunde oder während des Experimentes eine Frage haben, drücken Sie den roten Knopf auf Ihrer Tastatur (F11). Der Experimentleiter wird dann zu Ihnen kommen und Ihre Fragen privat beantworten.

C.3 Screenshots

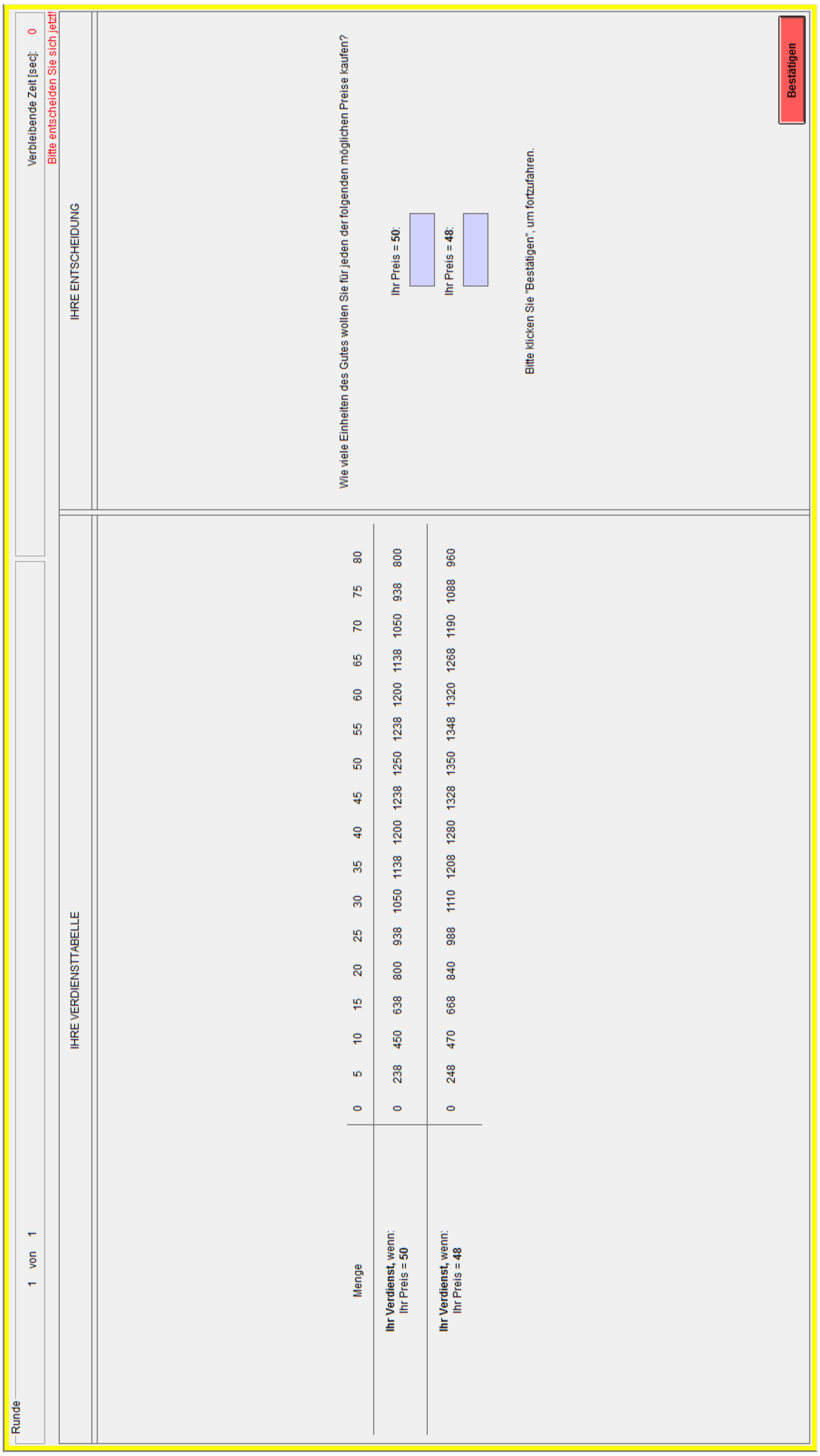


Figure C.1: Screen of consumer 1 in *No Info* treatment

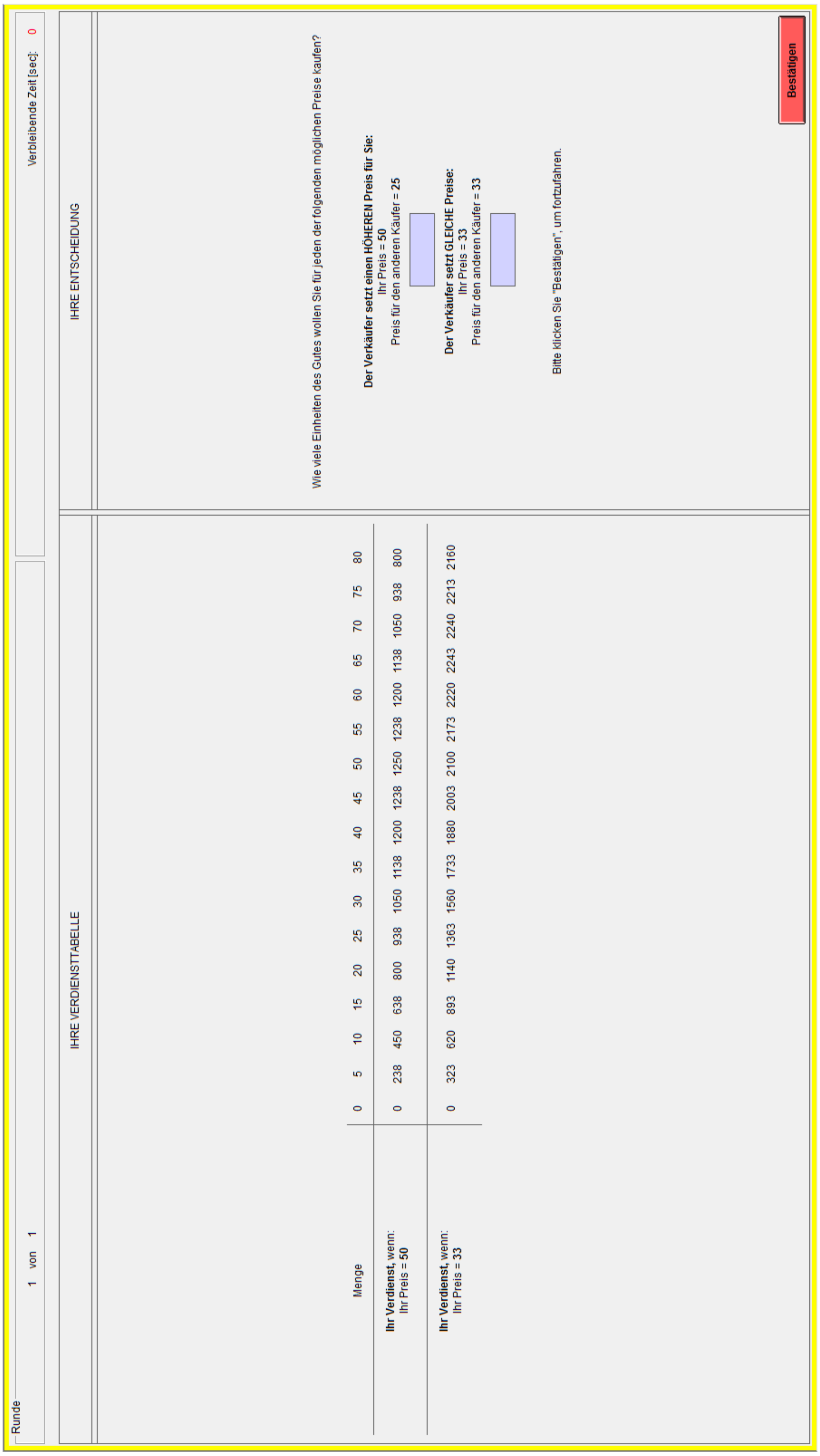


Figure C.2: Screen of consumer 1 in *Part Info* treatment

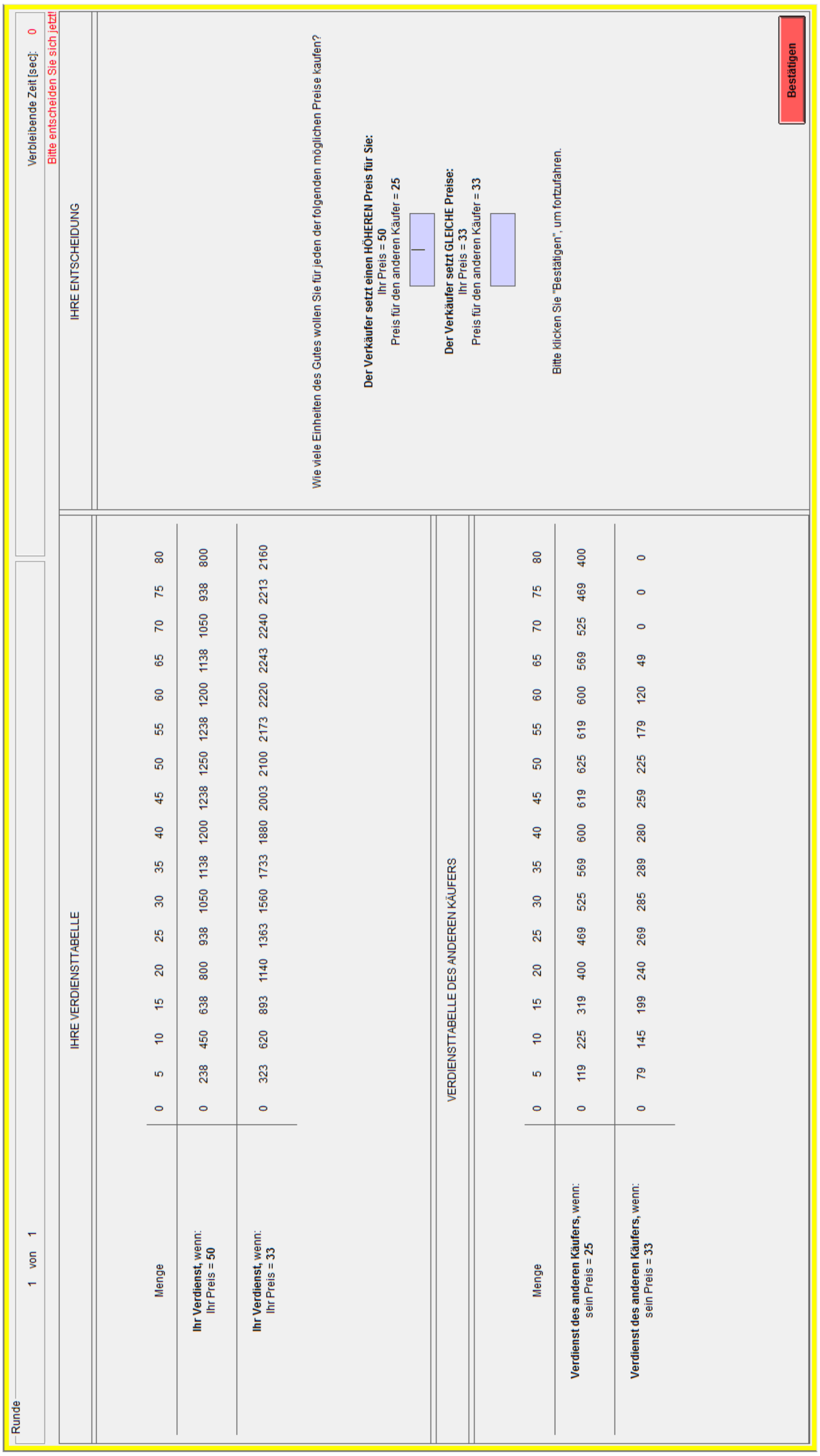


Figure C.3: Screen of consumer 1 in *Full Info* treatment

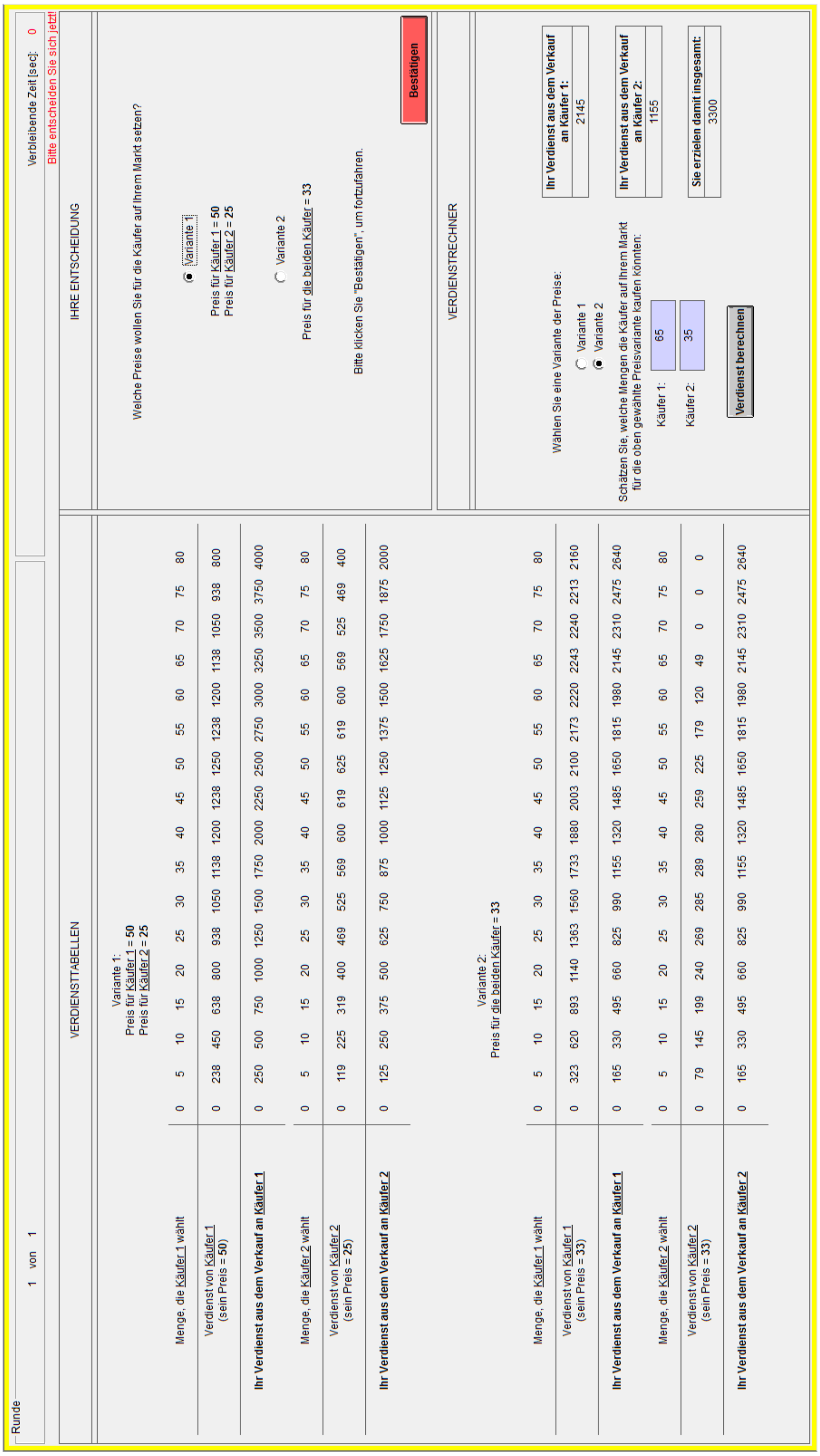


Figure C.4: Screen of the firm in all treatments

C.4 Likelihood function

The choices of the consumers in our experiment were discrete: consumers could pick quantities from $\{0, 5, 10, ..., 75, 80\}$. Following Cox et al. (2007), we calculate the threshold values of the error term at which optimal choices change. We define ϵ_i^q as the threshold where consumer i is indifferent between buying q and $q+5$, all else being equal:

$$U_i^\epsilon(q, p_i, p_j, \epsilon_i^q) = U_i^\epsilon(q+5, p_i, p_j, \epsilon_i^q).$$

Substitung for the utility functions and solving for ϵ_i^q yields the following formula:

$$\epsilon_i^q = \frac{u_i(q, p_i) - u_i(q+5, p_i)}{5p_i} - \rho_i \Delta_i(p_i, p_j, \gamma_i).$$

Thus, we derive 80 thresholds of the error term $\{\epsilon_i^0, \epsilon_i^5, \epsilon_i^{10}, ..., \epsilon_i^{75}\}$. For $\epsilon_i \in (-\infty, \epsilon_i^0)$, it is optimal for consumer i not to buy anything from the firm. For any $q \in \{5, 10, ..., 75\}$, consumer i would buy q if $\epsilon_i \in (\epsilon_i^{q-1}, \epsilon_i^q)$. Finally, consumer i prefers to buy 80 if $\epsilon_i \in (\epsilon_i^{75}, \infty)$. The probability of purchasing q can be determined using the cumulative distribution function of the normal distribution with mean α_i and variance σ_i^2:

$$Pr[q_i = q|p_i, p_j, \alpha_i, \sigma_i, \rho_i, \gamma_i] = \begin{cases} F\left(\epsilon_i^0 \,|\alpha_i, \sigma_i^2\right) & \text{if } q = 0; \\ F\left(\epsilon_i^q \,|\alpha_i, \sigma_i^2\right) - F\left(\epsilon_i^{q-1} \,|\alpha_i, \sigma_i^2\right) & \text{if } q \in \{5, 10, ..., 75\}; \\ 1 - F\left(\epsilon_i^{75} \,|\alpha_i, \sigma_i^2\right) & \text{if } q = 80. \end{cases} \tag{18}$$

These probabilities, in turn, determine the likelihood function:

$$L_i(p_i, p_j, \alpha_i, \sigma_i, \rho_i, \gamma_i) = \prod_{k=1}^{40} Pr[q_{ik} = q|p_i, p_j, \alpha_i, \sigma_i, \rho_i, \gamma_i].$$

Maximizing its logarithm, we arrive at the desired parameter estimates.

Appendix D The determinants of gender bias in student evaluations of teaching

D.1 Instructions

Below we provide the paper-based instructions for the subjects in the laboratory. This text was handed out in *PrizeInfo* treatments; in *NoPrizeInfo* treatments, the text highlighted in red was missing.

Welcome to the experiment and thank you for your participation!

Please do not talk to other participants from now on.

This experiment analyzes economic decision making. During the experiment you and the other participants will make decisions and you will earn money. The amount of money you earn depends only on your own decisions and is determined by the rules that will be explained in these instructions. At the end of the experiment your total profit will be paid to you privately in cash. You will additionally receive 6 Euros as a show-up fee.

The whole experiment will last about 60 minutes. If you have any questions after reading the instructions please raise your hand. One of the experimenters will come to you and answer your questions privately. All participants receive the same instructions.

The other participants will neither during nor after the experiment learn how much you earned. We never link names and data from experiments. At the end of the experiment you will be asked to sign a receipt regarding your earnings which serves only as a proof for our sponsor. The latter does not receive any other data from the experiment.

While you make your decisions, a clock at the top of your computer screen will run down. This clock will inform you regarding how much time you still have to make your decision.

The experiment consists of three parts. You will receive instructions for each part after the previous part has ended.

Part 1

In the first part of the experiment, you will watch a 20-minute videolecture on a mathematical topic and then take a test on its content. The test consists of 10 questions. You will earn 1 Euro for each correct answer, and nothing for incorrect answers. You can take up to 20 minutes to complete the test. Please press the button "SUBMIT" when you are ready. If you do not press "SUBMIT" before 20 minutes run out, you will not earn anything in this part of the experiment. You will learn how many of your answers are correct in the end of the experiment.

There is a pen and two sheets of paper on your table. Feel free to use them to take notes during the videolecture and the test if you wish to do so. One sheet is for your notes during the lecture, and the other one is for your notes during the test. We ask you to leave the pen and the paper on your table when the experiment ends. Your notes will not affect your earnings, will not be transferred to anyone and will be destroyed when our study is completed.

If you have any questions during the experiment, please raise your hand. We ask you to put on the headphones now. When the videolecture is over, you may take them off.

Part 2

We ask you to complete a short evaluation questionnaire on the lecture from Part 1. We hired 8 instructors to prepare the videolecture equivalent to the one that you saw. After all sessions of our experiment are completed, we will tell the instructors their personal average evaluation scores on each question. The average is taken over the evaluations of all participants who saw the respective lecture. Your evaluation thus stays anonymous.

We will give a prize of 50 Euros to the instructor that receives the highest average evaluations in our experiment.

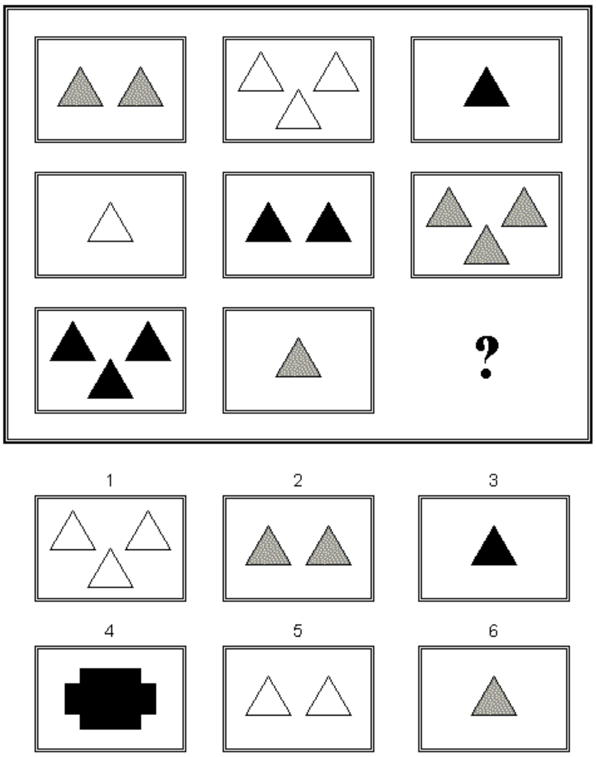
You have 5 minutes to complete the evaluation form.

Part 3

In this part, we ask you to solve a series of 8 problems. All have the same structure.

You see a box with a matrix, i.e. a rectangular pattern of different symbols. Each matrix has 3 rows and 3 columns. The symbol in the bottom right corner is missing.

Below the matrix, there are 8 symbols to choose from. Only one of them fits the pattern and should go into the bottom right corner of the matrix. Here is an example:



The correct solution is symbol number 5. The task for you is to choose the correct symbol. After you have done that for one problem, the next problem appears on the screen. You cannot go back to earlier problems, once you have submitted the solution.

For every correct solution, you will earn 50 Euro-cents. For every incorrect solution, you will earn nothing. You will have 60 seconds for every problem. If you do not submit any solution to a problem before the time runs out, it is assumed that the solution to this problem is incorrect.

After Part 3 is completed, we ask you to fill out a short questionnaire on the screen. Then you will learn how many problems you have solved correctly and how much you have earned. In addition, you will learn your score on the test and your corresponding earnings from Part 1. The sum of your earnings in Part 1 and Part 3 will be your total earnings from the experiment. Only you but no other participant will receive the information on your earnings. Then the experiment ends and we are going to call you for payments.

D.2 Lecture script

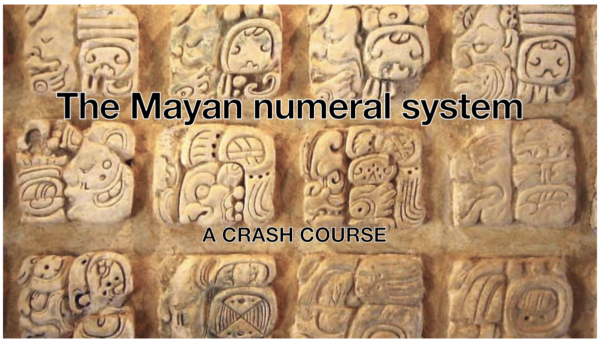
The text highlighted in blue and the slides in blue frames are present only in *HighQ* treatment. The text highlighted in red and the slides in red frames are present only in *LowQ* treatment. The rest of the script and the slides is identical in both treatments.

Video ON

Hello everyone! My name is %Johannes% and I am going to give you a lecture on the numeral system of the ancient Maya. This lecture will be followed by a short test, where you will solve 10 problems on Mayan numbers. You are welcome to use your pens and paper to make notes during the lecture and the test. At the end of this crash course you will become a pro in mathematics of the Maya civilization. So, let us start!

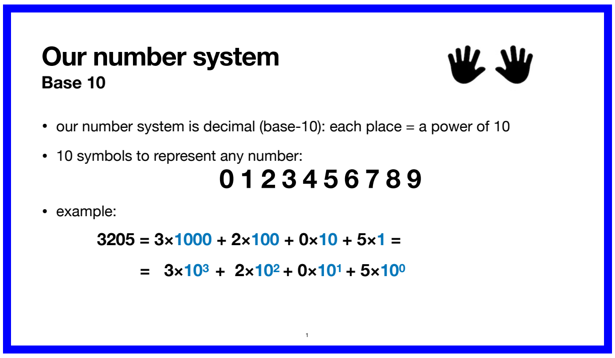
Video OFF: slides on the screen

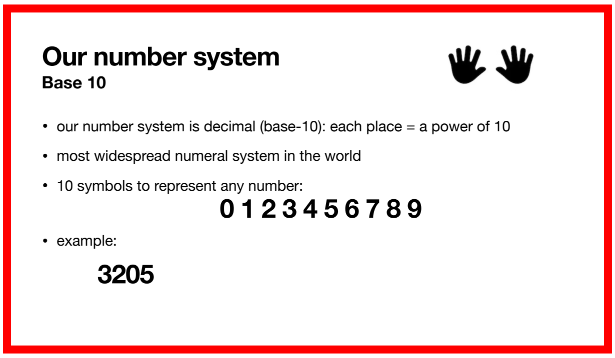
[the slides are animated in *HighQ* treatment and not animated in *LowQ* treatment]



In this lecture I will explain you the basics of the Mayan numeral system. We will start with the history, learn how Mayan numbers look like, compare them with our numbers and calculate some examples.

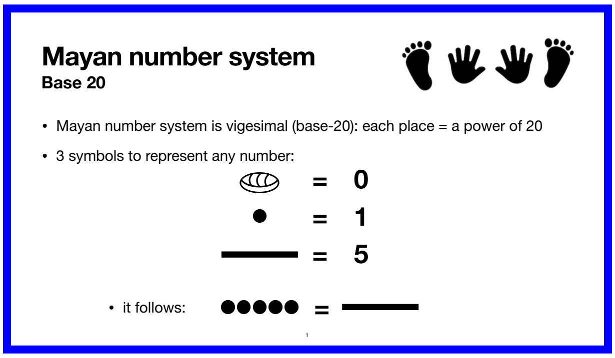
The Maya civilization started out in Central America thousands of years ago. It experienced the peak of its development from 250 A.D. to 900 A.D. During this time, the Mayas made impressive advancements in arts, culture, architecture, mathematics and astronomy. They developed an efficient numeral system of their own, possibly one of the world most advanced at the time. It allowed the Maya to do the elaborate calculations needed to make precise astronomical predictions. The precision of their observations along with their astronomical and calendrical record keeping were astonishingly accurate. One of the great intellectual accomplishment of the Maya was the use of the mathematical concept of zero. They were one of the only ancient civilisations to use it, and the appearance of the zero in Maya inscriptions is one of the earliest known instance of this concept in the world.

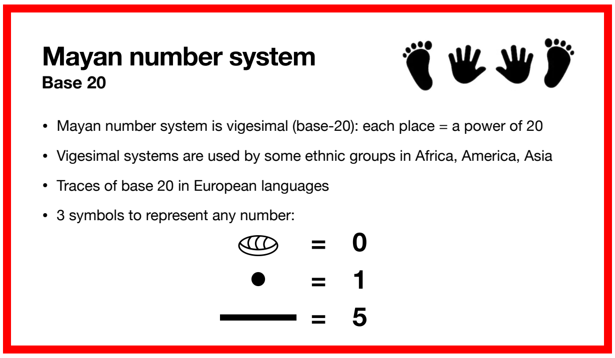




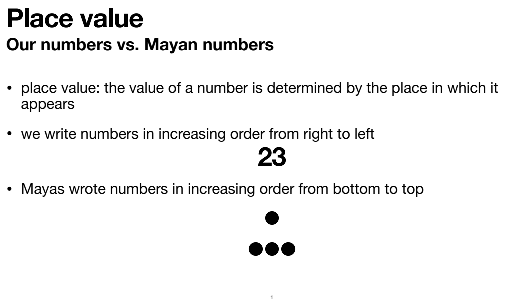
Before we dive into the world of Mayan numbers, let us have a closer look at the number system we use today. Our numeral system is decimal, or base ten. It means that the digit in each position is a power of 10. Our number system was designed as base 10 simply because 10 equals the number of our fingers. Many numeral systems of ancient civilizations use ten and its powers for representing numbers, including Brahmi numerals, Greek numerals, Hebrew numerals, Roman numerals, and Chinese numerals. The numeral system we use today stems from Hindu-Arabic system, which was invented between the 1st and 4th centuries by Indian mathematicians. Arabic numerals were introduced to Europe in the 10th century by Arabic-speakers of North Africa. The European acceptance of the numerals was accelerated by the invention of the printing press, and they became widely known during the 15th century. In this system we have 10 symbols (or 10 digits), with which we can write down any number. Let us have a look at the following example: the number 3205. The digit 3 denotes the number of thousands contained in this number. The digit 2 denotes the number of hundreds.

The digit 0 denotes the number of tens, and the digit 5 denotes the number of ones. A thousand, a hundred and ten are all powers of 10. One thousand is equal to 10 to the power of 3, one hundred is 10 to the power of 2, and ten is 10 to the power of 1. If we add up 3 thousands, 2 hundreds, 0 tens and 5 ones, we are going to get precisely 3205.

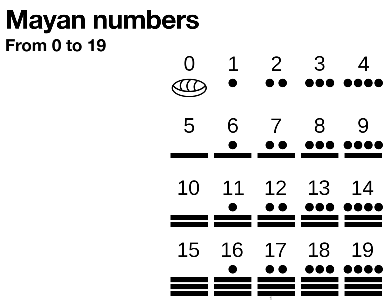




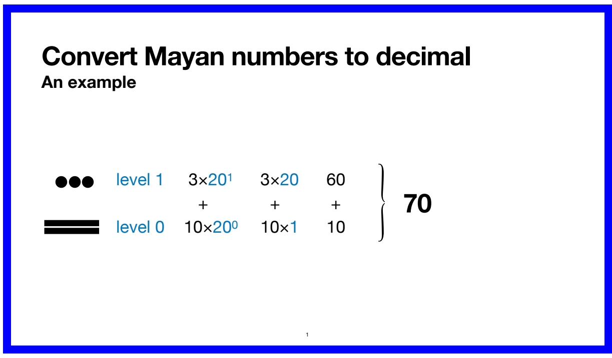
What about the Mayan numbers? Unlike us, the Maya used 20 as the base for their numeral system. Such a system is called vigesimal. Scholars believe that they chose 20 to be the base because it is the combined total of fingers and toes. Vigesimal systems are still used in different corners of the world: for example, among Yoruba people in Africa, Alaska Natives in America, Burushaski people in Pakistan and Chukchi people in Siberia. We can also find traces of base 20 in many European languages. For example, in English, Slavic languages and German the names of the two-digit numbers from 11 to 19 consist of one word, but the names of the two-digit numbers from 21 on consist of two words. In French, twenty is used as a base number in the names of numbers from 70 to 99. Twenty is also used as a base for some numbers in Danish, Albanian, Georgian, Welsh and Scottish Gaelic languages. Now let us return to the Mayan system. To represent any possible number, the Maya had only three symbols. The shape of a conch shell denotes 0; a dot denotes 1; and a horizontal bar denotes 5. Thus, every five dots are equal to one bar.

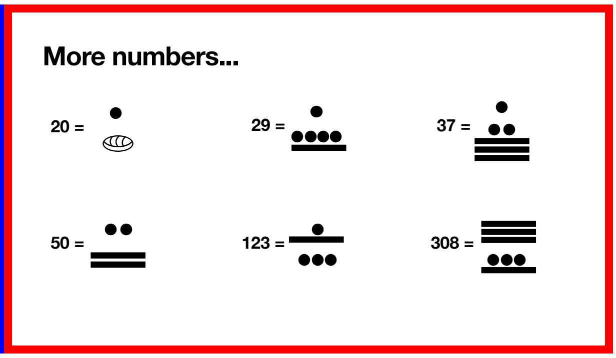


Another difference of the Mayan numeral system is the place value. When we write down numbers, the value of a digit is determined by the place in which it appears. In our decimal system, the place values are powers of 10, in increasing order from right to left. For example, look at the number 23. The digit 3 represents the number of ones, and the digit 2 to the left of it represents the number of tens. Thus, 2 has a higher place value as 3. The Mayas wrote their numbers in increasing order from the bottom to the top. Here you see the number 23 written in the Mayan system. The bottom row has 3 dots that correspond to 3 ones. One dot in the top row has a higher place value: namely, it denotes the number of 20-s. It might seem natural to you that the value of the digit depends on its position; however, there are numeral systems for which it is not the case. For instance, in Roman numbers a digit has only one value: I means one, V means five, X means ten and C a hundred (however, the value may be negated if placed before another digit). The problem with such systems is that when numbers get larger, new symbols must be invented. This makes writing large numbers difficult because so many new symbols need to be memorized. In contrast, the existence of place value in the Mayan numeral system allowed them to write down any number with the help of only three symbols.

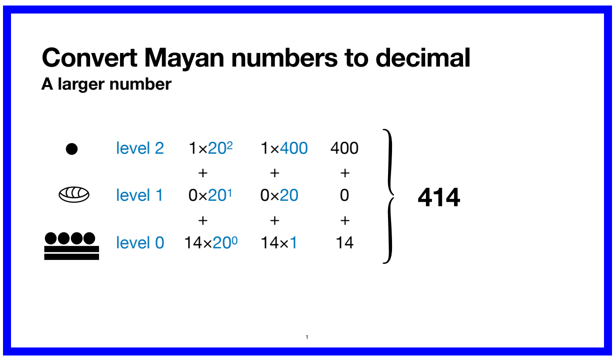


Now let us count from 0 to 19 using Mayan numerals. Zero is a shell, one is one dot, two is two dots, three is three dots, four are four dots, and then we move to five – five is a bar. To produce six, we throw a dot on the top of the bar. One bar and one dot equals to 5 plus 1, equals to 6. Adding another dot yields seven, then eight, then nine, and then we reach ten – which is two bars. Remember that two bars are two times 5, which equals 10. Putting additional dots on the top of these two bars, we get 11, 12, 13, and 14, respectively. Fifteen corresponds to three bars, or three times 5. By adding dots one the top of the three bars one by one, we get to 16, 17, 18 and 19.

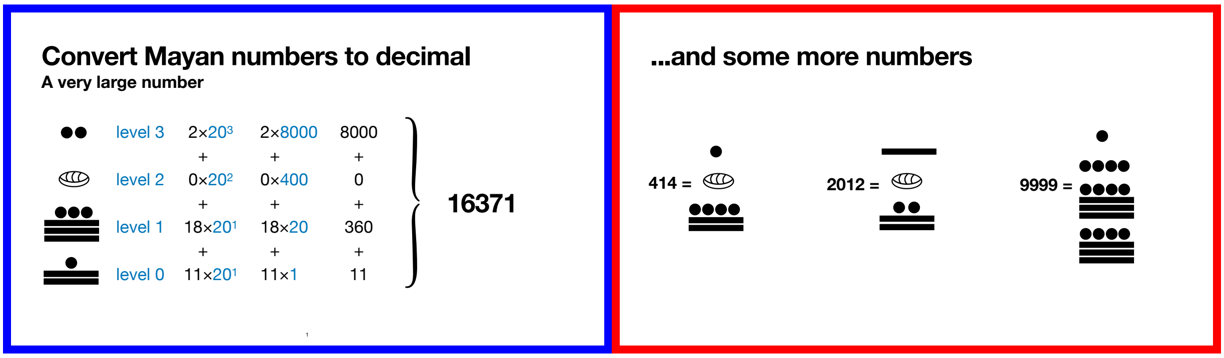




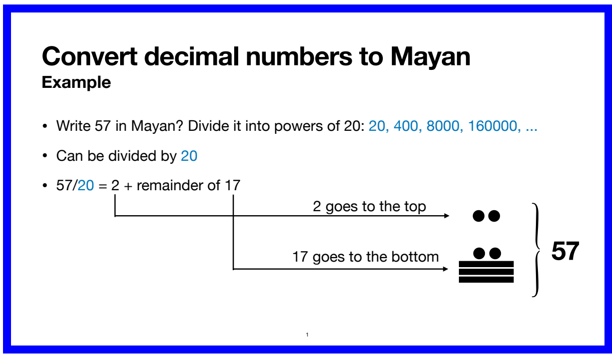
Now I want you to learn how to read larger numbers in Mayan numeral system. Let us consider an example of how to convert a Mayan number into the decimal system. Here we see 2 bars at the bottom level and 3 dots at the top level. I will call the bottom level "level 0" since its place value corresponds to ones, which are 20 to the power of 0; and the top level "level 1" since its place value corresponds to 20s, or 20 to the power of 1. Recall that two bars represent number 10; and three dots are number 3. So, we have 10 at level 0 and 3 at level 1, which amounts to 10 times 1 plus 3 times 20, which is equal to 10 plus 60, which is equal to 70. The number we saw on this picture is equal to 70. Let us look at the following examples. Here we see number 20 – a dot at the top corresponding to one base of 20 and a zero at the bottom level. Number 29 has the same dot above and the Mayan number 9 below. Number 37 also has one dot at the top since it can only fit one base of 20. The remainder of 17 is drawn at the bottom level: three bars with two dots. Now look at the number 50. It contains two 20s, that is why two dots are drawn on the top level. Apart from it, it contains a 10, which is drawn on the bottom level as two bars. Next to it you see a larger number – 123. Six 20s fit into this number, so we draw a bar with a dot on the top. Three dots are left as a remainder on the bottom. Finally, number 308: it contains 15 bases of 20, so three bars appear on the upper level. The remainder of 8, or a bar with 3 dots, goes onto the lower level.

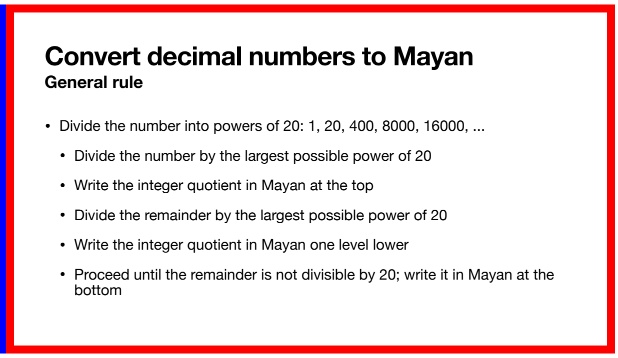


Here is an example of a larger number with more levels. At level 0 we have two bars and four dots, which is 14. At level 1 there is a zero, at level two there is one dot. It corresponds to 14 ones plus zero twenties plus one 400. 14 plus 0 plus 400 adds up to 414.

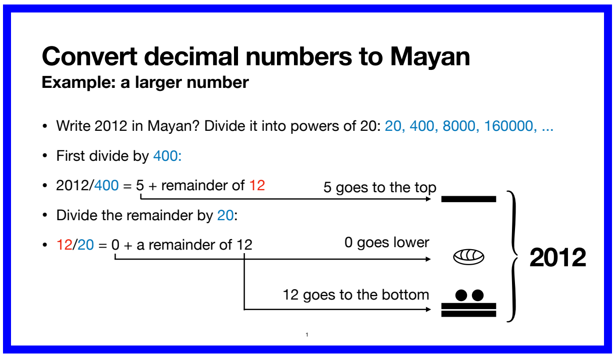


Finally, here is a very large number with four levels. Two bars and a dot at level 0 mean 11; three bars with three dots at level 1 mean 18. There is a zero at level 2 and two dots at level 3. It translates to 11 ones, 18 twenties, zero 400s, and two 8000s. 11 plus 360 plus 0 plus 16000 is a total of 16371. Here are some examples of larger numbers with more levels. Look at the number 414. At the lowest level we have two bars and four dots, which is 14. In the middle there is a zero, at the top there is one dot. It corresponds to 14 ones plus zero twenties plus one 400. The next number is 2012. It consists of 12 on the bottom level, zero 20-s in the middle and five 400s on the top. By the way, you might recall that the Maya are credited for predicting the end of the world in year 2012. Around 300 B.C., Mayan priests designed a calendar known as the Long Count, which identified each day by counting forward from a base point calculated to fall on August 11, 3114 B.C. A single cycle of the Long Count calendar lasts roughly 5126 solar years, meaning that it is slated to end on a date correlating to December 21, 2012. Popular culture has latched on to theories that the close of the calendar's current cycle corresponds to the end of the world in the Mayan belief system. However, no cataclysm happened and now you know how to write down year 2012 in the Mayan numeral system. Now let us look at the last example, a larger number: 9999. It has one dot on the top level, which corresponds to one 8000, or one 20 to the power of 3. 4 dots at the lower level denote four 400s. The Mayan 19 on the next lower level is the number of 20s, and a remainder of 19 is on the bottom level. Now you are able to understand even very large numbers in the Mayan system.

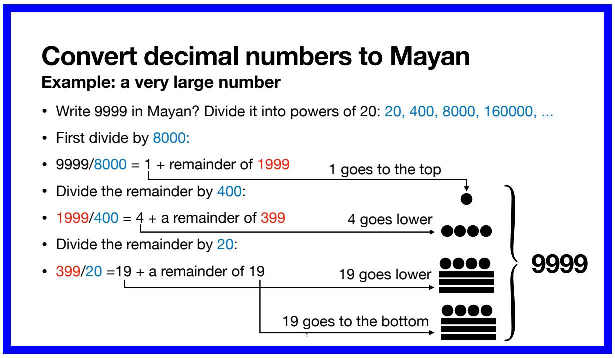




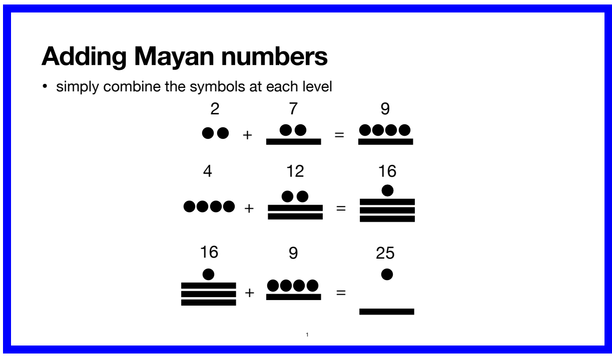
Now suppose that you want to solve a reverse problem: you have a number in the decimal system, and you want to convert it into the Mayan system. So, how would you write it down? To figure out the Maya equivalent of any number, you need to decompose it into the powers of 20, which are 20, 400, 8000, 160000 and so on. Let us start with a simple example and try to write down number 57. First, we look at it and ask: what is the maximum possible power of 20 that 57 can be divided with? In this case, 57 is a small number and it can only be divided by 20. In the next step, we divide 57 by 20, which gives us 2 and a remainder of 17. Getting 2 after division means that there are two 20s in our number 57; therefore we can draw two dots on the top level of the Mayan number. The remainder of 17 denotes 17 ones, and it goes to the bottom level of the Mayan number. Here we are: number 57 in Mayan consists of two 20s at the top and 17 ones at the bottom. First, we look at our number and ask: what is the maximum possible power of 20 that this number can be divided with? We divide our number with this largest possible power of 20 and write down the resulting integer quotient as a Mayan number at the top level. In the next step, we take the remainder of this division and divide it by a smaller power of 20. Again, the resulting integer quotient is written down at the level below. The remainder is again divided by another smaller power of 20, and so on. We repeat this procedure until the remainder is not divisible by 20, and write down this remainder at the very bottom. Now I am sure you can convert any number back and forth between Mayan and decimal numeral systems.

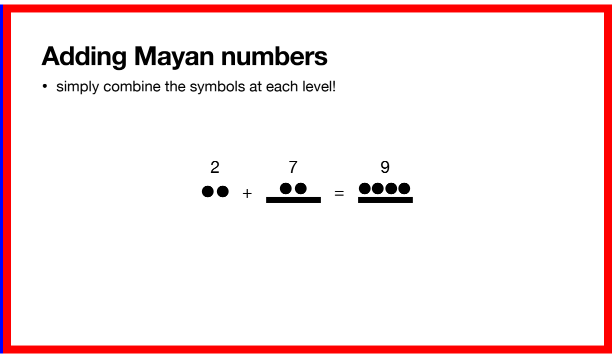


Let us have another example and convert number 2012 into Mayan system. Again, we look at the powers of 20 that could fit into 2012. In this case, 400 is the maximum possible divisor. We start by dividing 2012 with 400. It gives us 5 plus a remainder of 12, meaning that our number contains five 400s. Therefore, we can draw 5 or a horizontal bar at the top level. Now we go one level lower and divide the remainder of 12 by the next power of 20, which is 20 in our case. Since 12 is less than 20, the result of this division is 0 plus a remainder of 12. We write zero or a shell at the level below. Finally, the last remainder of 12 goes to the bottom level – two bars and two dots. This is the result: 2012 in Mayan consists of five 400s, no 20s and 12 ones.

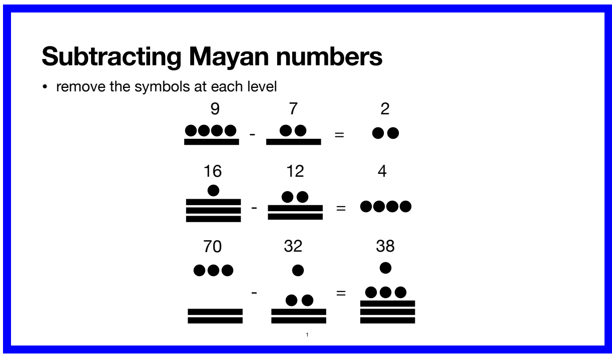


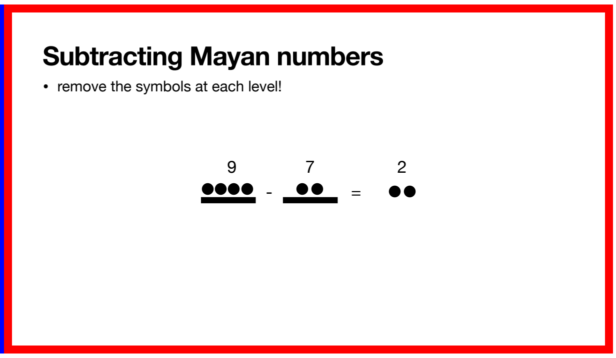
Finally, I want you to write down a very large number in Mayan: let us take 9999. First, which power of 20 can you divide with? Can you divide it by 20? Yes. 400? Yes. 8000? Also yes, but not more. So let us divide 9999 by 8000. It gives us 1 and a remainder of 1999. We draw one dot on the top. Then we move one level lower and divide the remainder of 1999 by the next power of 20, which is 400. 1999 divided by 400 is 4 plus a remainder of 399. We draw 4 dots one level below. Next step: we divide 399 by 20, and this is equal to 19 plus a remainder of 19. So we draw 19 one level below – 3 bars and 4 dots. And the remainder of 19 is left on the bottom – also 3 bars and 4 dots. 9999 in Mayan system is ready: one 8000, four 400s, 19 twenties and 19 ones. Now I am sure you can convert any number back and forth between Mayan and decimal numeral systems.





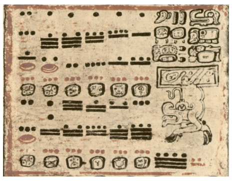
In the last part of this lecture, we will have a look at two simple operations with Mayan numbers: addition and subtraction. First, let us look at addition. To add any Mayan numbers, you simply put together the symbols at each level. Here is an example: two dots plus a bar with two dots. The sum is simply the bar at the bottom and all the dots combined at the top; or a bar with 4 dots. Ready! We performed addition without converting the Mayan numbers into the decimals. But if you want to, you can check: this is a 2, this is a 7, and the total is a 9. Another example: 4 dots plus two bars with 2 dots. When we add up, we pile the bars at the bottom, and throw all dots on them. This would leave us with 2 bars and 6 dots. However, remember that the Maya converted each 5 dots into a bar. So, the final result will be 3 bars with 1 dot. Here are the corresponding decimals: 4 plus 12 equals 16. The last example with more levels: let us add 3 bars with a dot and a bar with 4 dots. We pile all bars together and throw all dots on top; it leaves us with 4 bars and 5 dots. 5 dots turn into a bar, so now we have 5 bars. And now remember that every 4 bars constitute a 20: 4 times 5 is 20. We can wipe out 4 bars from this level and convert them into a dot on the upper level. The respective decimal numbers: 16 plus 9 equals 25. You have seen that Mayan addition was a relatively simple matter of adding up dots and bars. Now you might find it even more visual and intuitive than operations in our own numeral system. We have evidence of Maya's working with sums up to the hundreds of millions, and with dates so large it took several lines just to represent them. Despite not possessing the concept of a fraction, they produced extremely accurate astronomical observations using no instruments other than sticks and were able to measure the length of the solar year to a far higher degree of accuracy than that used in Europe. Their calculations were extremely close to the modern value of slightly more than 365 days.





Last but not least, let us subtract some Mayan numbers. Subtraction works in the same manner: you take away the symbols at each level. An example: a bar with 4 dots minus a bar with 2 dots. You take away the bar and you take away 2 dots; 2 dots are left. 9 minus 7 equals 2. Another example: 3 bars with 1 dot minus 2 bars with 2 dots. You take away the two bars and you can take away one dot easily, and now you have another dot to take away, but there are no dots left. Now you can convert one bar into 5 dots and take one dot away; it leaves you with 4 dots. 16 minus 12 equals 4. A final example with more levels: two bars on the bottom and three dots on the top minus two bars with two dots at the bottom and one dot on the top. Here you just do the subtraction at each level separately. Starting at the bottom, you wipe out two bars, and you need to take away two dots more, but nothing is left at this level. What you do is you take one dot from the top and convert it into 4 bars at the lower level – because 4 bars constitute a 20. Now you can convert one bar into 5 dots and take away the two dots, which leaves you with the Mayan 18 at the bottom. Now let us look at the top level. We started off with 3 dots, transferred one of them onto the lower level, and now we subtract one dot from the remaining two, which leaves us with one dot on the top. The subtraction is over. The first number was 70, the second number was 32, and the result is 38. If there are not enough dots in the first term, you can convert one bar into 5 dots. It will give you the necessary amount of dots you can subtract from. If you need to subtract larger Mayan numbers with more levels, you just do the subtraction at each level separately by wiping out the respective amount of bars and dots. In case there are not enough dots or bars at some level, you can simply transfer one dot from the level above and convert it into bars or dots at the level below.

The crash course is completed!
Ready to take the quiz?



Now the lecture on Mayan numeral system has come to an end. I thank you for your attention. Now you are able to read and write Mayan numbers and perform simple operations on them. When this video ends, you will start to take a test on Mayan numbers. I wish you good luck and hope that you enjoyed this crash course!

D.3 Evaluation questionnaire

To what extent do you agree or disagree with the following statements? Please select one answer per row.

	← strongly agree … strongly disagree →
The lecture is well-structured.	☐ ☐ ☐ ☐ ☐
I learned something useful and important.	☐ ☐ ☐ ☐ ☐
The lecture covered the content of the follow-on test.	☐ ☐ ☐ ☐ ☐

	← top grade … lowest grade →
Subjective grade for the lecture:	☐ ☐ ☐ ☐ ☐

	← strongly agree … strongly disagree →
The instructor presents the subject matter in an understandable way.	☐ ☐ ☐ ☐ ☐
The instructor appears to be competent.	☐ ☐ ☐ ☐ ☐
The instructor appears to be well-prepared.	☐ ☐ ☐ ☐ ☐
The instructor is motivated and enthusiastic about the topic.	☐ ☐ ☐ ☐ ☐

	← top grade … lowest grade →
Subjective grade for the instructor:	☐ ☐ ☐ ☐ ☐

D.4 Questionnaire at the end of the experiment

Before the experiment ends, please answer the following questions:

- Your gender
- Your age
- Your major
- Your grade for mathematics in the last school year
- How would you evaluate your general mood today?
- How would you evaluate your mood during the experiment?

Please also state to what extent you agree or disagree with the following statements. Your answers will not be known by your instructor. Your answers will not affect the evaluation of your instructor that you completed in Part 2 of the experiment.

- I enjoyed the lecture.
- The instructor was confident.
- The instructor was humorous.
- The instructor was friendly.
- The instructor was charismatic.

Now imagine that you look back at this lecture 5 years from now.

- This lecture substantially improved my skills and/or knowledge.
- This lecture changed my way of thinking.
- This lecture helped me approach important problems in my life.

Bibliography

Abbink, K., Irlenbusch, B., Pezanis-Christou, P., Rockenbach, B., Sadrieh, A., and Selten, R. (2005). An experimental test of design alternatives for the British 3G/UMTS auction. *European Economic Review*, 49(2):505–530.

Al-Ubaydli, O., Gneezy, U., Lee, M. S., and List, J. A. (2010). Towards an understanding of the relative strengths of positive and negative reciprocity. *Judgment and Decision Making*, 5(7):524–539.

Allender, W., Liaukonyte, J., Nasser, S., and Richards, T. J. (2020). Price fairness and strategic obfuscation. Available at SSRN: `https://ssrn.com/abstract=2780170`.

Ambady, N. and Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology*, 64(3):431–441.

Amin, M., Hanson, K., and Mills, A. (2004). Price discrimination in obstetric services – a case study in Bangladesh. *Health Economics*, 13(6):597–604.

Anderson, E. T. and Simester, D. I. (2008). Does demand fall when customers perceive that prices are unfair? The case of premium pricing for large sizes. *Marketing Science*, 27(3):492–500.

Anderson, E. T. and Simester, D. I. (2010). Price stickiness and customer antagonism. *The Quarterly Journal of Economics*, 125(2):729–765.

Arbuckle, J. and Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles*, 49(9-10):507–516.

Arceo-Gomez, E. O. and Campos-Vazquez, R. M. (2019). Gender stereotypes: The case of MisProfesores.com in Mexico. *Economics of Education Review*, 72:55–65.

Armstrong, M. (2006). Recent developments in the economics of price discrimination. In Blundell, R., Newey, W. K., and Persson, T., editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress: Volume II*, pages 97–141. Cambridge University Press, Cambridge, UK.

Ashraf, N., Camerer, C. F., and Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives*, 19(3):131–145.

Ayres, I. and Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, 85(3):304–321.

Babad, E., Avni-Babad, D., and Rosenthal, R. (2003). Teachers' brief nonverbal behaviors in defined instructional situations can predict students' evaluations. *Journal of Educational Psychology*, 95(3):553–562.

Baker, R. J., Laury, S. K., and Williams, A. W. (2008). Comparing small-group and individual behavior in lottery-choice experiments. *Southern Economic Journal*, 75(2):367–382.

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4):323–370.

Berg, N. and Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, pages 133–165.

Bertrand, M. and Duflo, E. (2017). Field experiments on discrimination. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, volume 1, pages 309–393. Amsterdam: North-Holland.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Blinder, A. and Morgan, J. (2005). Are two heads better than one? Monetary policy by committee. *Journal of Money, Credit and Banking*, 37(5):789–811.

Bohren, J. A., Imas, A., and Rosenberg, M. (2019). The dynamics of discrimination: Theory and evidence. *American Economic Review*, 109(10):3395–3436.

Bolton, L. E., Warlop, L., and Alba, J. W. (2003). Consumer perceptions of price (un)fairness. *Journal of Consumer Research*, 29(4):474–491.

Borenstein, S. (1991). Selling costs and switching costs: Explaining retail gasoline margins. *RAND Journal of Economics*, 22(3):354–369.

Börgers, T. and Dustmann, C. (2005). Strange bids: Bidding behaviour in the United Kingdom's third generation spectrum auction. *The Economic Journal*, 115(505):551–578.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27–41.

Boring, A. and Philippe, A. (2017). Reducing discrimination through norms or information: Evidence from a field experiment on student evaluations of teaching. IAST working papers 17–72, Institute for Advanced Study in Toulouse.

Bornstein, G., Kugler, T., and Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game: Are groups more "rational" players? *Journal of Experimental Social Psychology*, 40(5):599–605.

Bougie, R., Pieters, R., and Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, 31(4):377–393.

Braga, M., Paccagnella, M., and Pellizzari, M. (2014). Evaluating students? Evaluations of professors. *Economics of Education Review*, 41:71–88.

Breaban, A. and Noussair, C. N. (2018). Emotional state and market behavior. *Review of Finance*, 22(1):279–309.

Burnell, N., Cojuharenco, I., and Murad, Z. (2020). He taught, she taught: The effect of teaching style, academic credentials, bias awareness and academic discipline on gender bias in teaching evaluations. Working papers in Economics & Finance 2020–05, University of Portsmouth, Portsmouth Business School, Economics and Finance Subject Group.

Buurman, M., Delfgaauw, J., Dur, R., and Zoutenbier, R. (2018). The effects of student feedback to teachers: Evidence from a field experiment. Tinbergen Institute Discussion Paper 2018-042/VII, Tinbergen Institute.

Campbell, M. C. (1999). "Why did you do that?" The important role of inferred motive in perceptions of price fairness. *The Journal of Product & Brand Management*, 8(2):145–153.

Campbell, M. C. (2007). "Says who?!" How the source of price information and affect influence perceived price (un)fairness. *Journal of Marketing Research*, 44(2):261–271.

Campos-Mercade, P. (2018). Power analysis through simulations in Stata: A guide for dummies. Unpublished manuscript. Accessed on 18-08-2020 at `https://sites.google.com/site/polcamposmercade/resources`.

Capen, E. C., Clapp, R. V., and Campbell, W. M. (1971). Competitive bidding in high-risk situations. *Journal of Petroleum Technology*, 23(06):641–653.

Carrell, S. E. and West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3):409–432.

Casari, M., Zhang, J., and Jackson, C. (2016). Same process, different outcomes: Group performance in an acquiring a company experiment. *Experimental Economics*, 19(4):764–791.

Charness, G. and Sutter, M. (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives*, 26(3):157–76.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5):1–33.

Chisadza, C., Nicholls, N., and Yitbarek, E. (2019). Race and gender biases in student evaluations of teachers. *Economics Letters*, 179:66–71.

Clayson, D. E. and Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, 28(2):149–160.

Cooper, D. J. and Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, 95(3):477–509.

Courty, P. and Pagliero, M. (2010). Price variation antagonism and firm pricing policies. *Journal of Economic Behavior & Organization*, 75(2):235–249.

Cox, J. C., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 1(59):17–45.

Cox, J. C. and Hayne, S. C. (2006). Barking up the right tree: Are small groups rational agents? *Experimental Economics*, 9(3):209–222.

Darke, P. R. and Dahl, D. W. (2003). Fairness and discounts: The subjective value of a bargain. *Journal of Consumer psychology*, 13(3):328–338.

Destan, C. G. and Yılmaz, M. (2020). Nonlinear pricing under inequity aversion. *Journal of Economic Behavior & Organization*, 169:223–244.

Doubleday, A. F. and Lee, L. M. (2016). Dissecting the voice: Health professions students' perceptions of instructor age and gender in an online environment and the impact on evaluations for faculty. *Anatomical Sciences Education*, 9(6):537–544.

Doyle, L. and Schindler, D. (2019). μCap: connecting FaceReader™ to z-Tree. *Journal of the Economic Science Association*, 5(1):136–141.

Dubé, J.-P. and Misra, S. (2017). Scalable price targeting. NBER Working Paper No. w23775, National Bureau of Economic Research.

Englmaier, F., Gratz, L., and Reisinger, M. (2012). Price discrimination and fairness concerns. Available at SSRN: `https://ssrn.com/abstract=1999757`.

Eyméoud, J.-B. and Vertier, P. (2017). Gender biases: Evidence from a natural experiment in French local elections. Available at SSRN: `https://ssrn.com/abstract=2913453`.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Fan, Y., Shepherd, L., Slavich, E., Waters, D., Stone, M., Abel, R., and Johnston, E. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE*, 14(2):e0209749.

Fehr, E. and Fischbacher, U. (2002). Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478):1–33.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Feinberg, F. M., Krishna, A., and Zhang, Z. J. (2002). Do we care what others get? A behaviorist approach to targeted promotions. *Journal of Marketing Research*, 39(3):277–291.

Feri, F., Irlenbusch, B., and Sutter, M. (2010). Efficiency gains from team-based coordination – large-scale experimental evidence. *American Economic Review*, 100(4):1892–1912.

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

Frey, B. S. and Pommerehne, W. W. (1993). On the fairness of pricing – an empirical survey among the general population. *Journal of Economic Behavior & Organization*, 20(3):295–307.

Funk, P., Iriberri, N., and Savio, G. (2019). Does scarcity of female instructors create demand for diversity among students? Evidence from observational and experimental data. Available at SSRN: `https://ssrn.com/abstract=3504620`.

Garbarino, E. and Maxwell, S. (2010). Consumer response to norm-breaking pricing events in e-commerce. *Journal of Business Research*, 63(9-10):1066–1072.

Gielissen, R., Dutilh, C. E., and Graafland, J. J. (2008). Perceptions of price fairness: An empirical research. *Business & Society*, 47(3):370–389.

Gigerenzer, G. (2018). The bias bias in behavioral economics. *Review of Behavioral Economics*, 5(3-4):303–336.

Gillet, J., Schram, A., and Sonnemans, J. (2009). The tragedy of the commons revisited: The importance of group decision-making. *Journal of Public Economics*, 93(5-6):785–797.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4):715–741.

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

Grewal, D., Hardesty, D. M., and Iyer, G. R. (2004). The effects of buyer identification and purchase timing on consumers' perceptions of trust, price fairness, and repurchase intentions. *Journal of Interactive Marketing*, 18(4):87–100.

Habel, J., Schons, L. M., Alavi, S., and Wieseke, J. (2016). Warm glow or extra charge? The ambivalent effect of corporate social responsibility activities on customers' perceived price fairness. *Journal of Marketing*, 80(1):84–105.

Haws, K. L. and Bearden, W. O. (2006). Dynamic pricing and consumer fairness perceptions. *Journal of Consumer Research*, 33(3):304–311.

Herz, H. and Taubinsky, D. (2018). What makes a price fair? An experimental study of transaction experience and endogenous fairness views. *Journal of the European Economic Association*, 16(2):316–352.

Ho, T.-H. and Su, X. (2009). Peer-induced fairness in games. *American Economic Review*, 99(5):2022–2049.

Hoffman, E., Marsden, J. R., and Saidi, R. (1991). Are joint bidding and competitive common value auction markets compatible? Some evidence from offshore oil auctions. *Journal of Environmental Economics and Management*, 20(2):99–112.

Huang, J.-H., Chang, C.-T., and Chen, C. Y.-H. (2005). Perceived fairness of pricing on the Internet. *Journal of Economic Psychology*, 26(3):343–361.

Huppertz, J. W., Arenson, S. J., and Evans, R. H. (1978). An application of equity theory to buyer-seller exchange situations. *Journal of Marketing Research*, 15(2):250–260.

Jagsi, R., Guancial, E. A., Worobey, C. C., Henault, L. E., Chang, Y., Starr, R., Tarbell, N. J., and Hylek, E. M. (2006). The"gender gap" in authorship of academic medical literature – a 35-year perspective. *New England Journal of Medicine*, 355(3):281–287.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Houghton Mifflin, Boston.

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. Houghton Mifflin, Boston.

Janis, I. L. (1989). *Crucial decisions: Leadership in policymaking and crisis management*. Free Press, New York.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986a). Fairness and the assumptions of economics. *The Journal of Business*, 59(4):285–300.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986b). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 76(4):728–741.

Kocher, M., Strauß, S., and Sutter, M. (2006). Individual or team decision-making – causes and consequences of self-selection. *Games and Economic Behavior*, 56(2):259–270.

Kocher, M. G., Schudy, S., and Spantig, L. (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9):3995–4008.

Kocher, M. G. and Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal*, 115(500):200–223.

Kugler, T., Kausel, E. E., and Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(4):471–482.

Lahey, J. N. and Oxley, D. R. (2018). Discrimination at the intersection of age, race, and gender: Evidence from a lab-in-the-field experiment. NBER Working Paper No. w25357, National Bureau of Economic Research.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, 504(7479):211–213.

Leibbrandt, A. (2020). Behavioral constraints on price discrimination: Experimental evidence on pricing and customer antagonism. *European Economic Review*, 121:103303.

Leslie, P. (2004). Price discrimination in Broadway theater. *RAND Journal of Economics*, 35(3):520–541.

Levine, D. K. (2012). *Is behavioral economics doomed? The ordinary versus the extraordinary*. Open Book Publishers, Cambridge.

Ley, T. J. and Hamilton, B. H. (2008). The gender gap in NIH grant applications. *Science*, 322(5907):1472–1474.

Li, K. J. and Jain, S. (2016). Behavior-based pricing: An analysis of the impact of peer-induced fairness. *Management Science*, 62(9):2705–2721.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, 119(1):49–89.

List, J. A. (2009). Social preferences: Some thoughts from the field. *Annual Review of Economics*, 1(1):563–579.

Maciejovsky, B., Sutter, M., Budescu, D. V., and Bernau, P. (2013). Teams make you smarter: How exposure to teams improves individual decisions in probability and reasoning tasks. *Management Science*, 59(6):1255–1270.

MacNell, L., Driscoll, A., and Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291–303.

Masclet, D., Colombier, N., Denant-Boemont, L., and Loheac, Y. (2009). Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers. *Journal of Economic Behavior & Organization*, 70(3):470–484.

Mengel, F., Sauermann, J., and Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2):535–566.

Moscovici, S. and Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12(2):125.

Naftulin, D. H., Ware, J. E., and Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48(7):630–635.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

Neumark, D., Bank, R. J., and Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: An audit study. *The Quarterly Journal of Economics*, 111(3):915–941.

Nguyen, Y. and Noussair, C. N. (2014). Risk aversion and emotions. *Pacific Economic Review*, 19(3):296–312.

Okada, T. (2014). Third-degree price discrimination with fairness-concerned consumers. *The Manchester School*, 82(6):701–715.

Ordóñez, L. D., Connolly, T., and Coughlan, R. (2000). Multiple reference points in satisfaction and fairness assessment. *Journal of Behavioral Decision Making*, 13(3):329–344.

Pech, W. and Milan, M. (2009). Behavioral economics and the economics of Keynes. *The Journal of Socio-Economics*, 38(6):891–902.

Peterson, D. A., Biederman, L. A., Andersen, D., Ditonto, T. M., and Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE*, 14(5):e0216241.

Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.com. *Journal of Diversity in Higher Education*, 3(3):137.

Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12):4403–4408.

Riach, P. A. and Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, 112(483):480–518.

Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. IZA Discussion Paper No. 8584, Institute for the Study of Labor, Bonn.

Rosen, A. S. (2018). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: A large-scale study of RateMyProfessors.com data. *Assessment & Evaluation in Higher Education*, 43(1):31–44.

Rotemberg, J. J. (2011). Fair pricing. *Journal of the European Economic Association*, 9(5):952–981.

Santos, G. and Dang Van Phu, S. (2019). Gender and academic rank in the UK. *Sustainability*, 11(11):3171.

Selove, M. (2019). Dynamic pricing with fairness concerns and a capacity constraint. *Quantitative Marketing and Economics*, 17(4):385–413.

Sheremeta, R. M. and Zhang, J. (2010). Can groups solve the problem of over-bidding in contests? *Social Choice and Welfare*, 35(2):175–197.

Shiller, B. R. (2020). Approximating purchase propensities and reservation prices from broad consumer tracking. *International Economic Review*, 61(2):847–870.

Shor, M. and Oliver, R. L. (2006). Price discrimination through online couponing: Impact on likelihood of purchase and profitability. *Journal of Economic Psychology*, 27(3):423–440.

Shupp, R. S. and Williams, A. W. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, 118(525):258–283.

Sonnabend, H. (2016). Fairness constraints on profit-seeking: Evidence from the German club concert industry. *Journal of Cultural Economics*, 40(4):529–545.

Stoner, J. A. F. (1961). *A comparison of individual and group decisions involving risk*. PhD thesis, Massachusetts Institute of Technology.

Sutter, M. (2005). Are four heads better than two? An experimental beauty-contest game with teams of different size. *Economics Letters*, 88(1):41–46.

Sutter, M., Kocher, M. G., and Strauß, S. (2009). Individuals and teams in auctions. *Oxford Economic Papers*, 61(2):380–394.

Thaler, R. H. (1997). Irving Fisher: Modern behavioral economist. *The American Economic Review*, 87(2):439–441.

Tsai, D. and Hsiao-Ching, L. (2007). Will you care when you pay more? The negative side of targeted promotions. *The Journal of Product and Brand Management*, 16(7):481–491.

Urbany, J. E., Madden, T. J., and Dickson, P. R. (1989). All's not fair in pricing: An initial look at the dual entitlement principle. *Marketing Letters*, 1(1):17–25.

van der Lee, R. and Ellemers, N. (2015). Gender contributes to personal research funding success in the Netherlands. *Proceedings of the National Academy of Sciences*, 112(40):12349–12353.

Verboven, F. (1996). International price discrimination in the European car market. *RAND Journal of Economics*, 27(2):240–268.

Wagner, N., Rieger, M., and Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review*, 54:79–94.

Wang, L. and Gonzalez, J. A. (2020). Racial/ethnic and national origin bias in SET. *International Journal of Organizational Analysis*, 28(4):843–855.

Wang, Y. and Krishna, A. (2012). Enticing for me but unfair to her: Can targeted pricing evoke socially conscious behavior? *Journal of Consumer Psychology*, 22(3):433–442.

Ware, J. E. and Williams, R. G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education*, 50(2):149–156.

Weisshaar, K. (2017). Publish and perish? An assessment of gender gaps in promotion to tenure in academia. *Social Forces*, 96(2):529–560.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., and Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE*, 8(7):e66212.

Wu, C.-C., Liu, Y.-F., Chen, Y.-J., and Wang, C.-J. (2012). Consumer responses to price discrimination: Discriminating bases, inequality status, and information disclosure timing influences. *Journal of Business Research*, 65(1):106–116.

Xia, L., Monroe, K. B., and Cox, J. L. (2004). The price is unfair! A conceptual framework of price fairness perceptions. *Journal of Marketing*, 68(4):1–15.

Zhang, J. and Casari, M. (2012). How groups reach agreement in risky choices: An experiment. *Economic Inquiry*, 50(2):502–515.