Implementation in the Presence of Social Preferences: A Behavioral and Experimental Economic Perspective

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2018

vorgelegt von

Lars Freund

aus

Bad Oldesloe

Referent:	Prof. Dr. Felix Bierbrauer
Koreferent:	Prof. Dr. Christoph Engel
Tag der Promotion:	30.04.2019

Acknowledgements

I am very grateful to my supervisors Felix Bierbrauer and Christoph Engel for their guidance and support during my whole time as a doctoral student. I belief that only the combination of your supervisions made it possible to write such a dissertation. I also owe special thanks to Nick Netzer who hosted and advised me greatly during my stay at the University of Zurich. I thank my co-author Franziska Tausch for her mentoring in the beginning of my time as a PhD candidate.

I would like to thank all my colleagues at the Max Planck Institute for Research on Collective Goods for the great time and making it feel like my second home. I am in dept for all their support, advise and constructive critics, in particular by Amalia Álverez, Philip Brookins, Minou Ghaffari, Adrian Hillenbrand, Svenja Hippel, André Schmelzer and Fabian Winter. I am also thankful to my office mates, Maj-Britt Sterba, Martin Sternberg and Eugenio Verrina who were an incredible source of motivation.

The administrative work to conduct a PhD should not be underestimated and I am specially grateful to Heidi Morgenstern, Doreen Hachenberg and Marion Schmitz for always being a great first contact point, being lenient with me and their support for all my wishes. The support by my family with respect to all aspects was amazingly helpful during my time as a student. Living in a scientific mind set limits from the time to time the point of view and I am thankful that I had my family, Marie-Christine Fries, Ximeng Fang, Nadja Kernchen, Stephan Martin, Jan Wergula and Anika Zadruzynski to remind me occasionally of the other aspects in life.

Contents

1 Introduction **1**

1.1 Environments . . . 2

1.2 Methods . . . 3

1.2.1 Game Theory and Mechanism Design . . . 3

1.2.2 Behavioral Economics . . . 4

1.2.3 Experimental Economics . . . 6

1.2.4 Mechanism Design and Social Preferences . . . 7

1.3 Chapter summaries . . . 8

2 Compulsory vs Voluntary Insurance **11**

2.1 Introduction . . . 11

2.2 Theoretical Framework . . . 13

2.2.1 Standard Preferences: $\gamma_i = 0$ & $\eta_i = 0$ . . . 14

2.2.2 Lying Aversion: $\gamma_i > 0$ & $\eta_i > 0$ . . . 15

2.2.3 Factors Influencing γ_i & η_i . . . 17

2.3 Experimental Design . . . 18

2.3.1 Risk Elicitation . . . 18

2.3.2 Main Experiment . . . 19

2.3.3 Questionnaire . . . 20

2.3.4 Experimental Procedure . . . 21

2.4 Results . . . 21

2.4.1 Insurance Purchase . . . 21

2.4.2 Cheating . . . 22

2.4.3 Insurer Profitability . . . 26

2.5 Discussion and Conclusion . . . 27

2.A Instructions . . . 29

2.B Questionnaire . . . 33

2.C Additional Informations . . . 35

2.C.1 Insurance Purchase Prediction . . . 35

2.C.2 Loss by Treatment . . . 36

3 Intention in Bilateral Trade 37

3.1 Introduction . . . 37
3.2 Theoretical Background . . . 40
3.2.1 Environment . . . 40
3.2.2 Analysis . . . 42
3.3 Experimental Design . . . 47
3.3.1 Online Experiment . . . 47
3.3.2 Laboratory Experiment . . . 49
3.3.3 Experimental Procedure . . . 51
3.4 Experimental Results . . . 52
3.5 Discussion . . . 61
3.6 Conclusion . . . 64
3.A Equilibrium Analyses of the Games . . . 66
3.B Equilibrium Figures . . . 68
3.C Subjects Behavior and Beliefs . . . 72
3.D Voluntarily Participation . . . 76
3.E Simulation for Subjects' Understanding . . . 77
3.F Questionnaire Detail . . . 78
3.G Experimental Instructions . . . 79

4 The Dependence of CM on Selfishness 85

4.1 Introduction . . . 85
4.2 Framework . . . 87
4.2.1 The Environment . . . 87
4.2.2 The Functional Form of the Utility . . . 88
4.2.3 The Equilibrium Concepts . . . 89
4.3 The Crémer-McLean Auction . . . 91
4.4 Behavioral Predictions . . . 96
4.5 Conclusion . . . 103
4.A Derivation of the Extraction Lottery . . . 104
4.B Ex Post Incentive Constraints . . . 104
4.C Ex Post Utility Function . . . 107

5 SRA: The Case of Correlated Valuations 109

5.1 Introduction . . . 109
5.2 Framework . . . 111

5.2.1 The Environment 111
5.2.2 Requirements on the Optimal Auction 112
5.2.3 The Crémer&McLean and Bartling&Netzer Auctions 113
5.3 Analysis 115
5.3.1 Optimal Auction for Bayesian Incentive Compatibility 117
5.3.2 Optimal Auction under Ex Post Incentive Compatibility . . . 122
5.4 Conclusion 125
5.A Expected profit maximization under BIC 127
5.B Expected profit maximization under EPIC 130

List of Figures

2.1 Ex post utility with lying aversion . . . 16
2.2 Cheating rate by treatment . . . 22
2.3 Cheating rate by treatment . . . 23
2.4 Predicted insurance purchase between treatments . . . 25
2.5 Insurer loss by treatment . . . 27
2.6 The actual losses subjects draw in each treatment . . . 36

3.1 Predicted effect of reciprocity on the likelihood of truth-telling . . . 45
3.2 Mini ultimatum game . . . 48
3.3 Treatment and hypotheses overview . . . 51
3.4 Frequency of the truth-telling strategy for the direct and indirect mechanism . . . 53
3.5 Kindness rating of the truth-telling strategy by the other subject . . . 54
3.6 First and second order beliefs that the other subject reports truthful . 54
3.7 Comparison of efficient trade in BASE and MSE: complete distribution & full efficiency . . . 55
3.8 Difference in beliefs that the other subject truthfully reveals . . . 58
3.9 Difference in truthful revelation of the private information . . . 59
3.10 Difference in efficient allocations . . . 60
3.11 Difference in happiness in the truth-telling strategy . . . 61
3.12 Difference in happiness in BASE and PM . . . 61
3.13 Equilibria in the BASE treatment: seller's point of view . . . 69
3.14 Equilibria in the BASE treatment: buyer's point of view . . . 70
3.15 Equilibria in the MSE treatment: seller's point of view . . . 70
3.16 Equilibria in the MSE treatment: buyer's point of view . . . 71
3.17 Equilibria in the PM treatment: seller's point of view . . . 71
3.18 Equilibria in the PM treatment: buyer's point of view . . . 72
3.19 Sellers' behavior and corresponding beliefs in BASE . . . 73
3.20 Buyers' behavior and corresponding beliefs in BASE . . . 73

3.21 Sellers' behavior and corresponding beliefs in MSE 74
3.22 Buyers' behavior and corresponding beliefs in MSE 74
3.23 Sellers' behavior and corresponding beliefs in PM 75
3.24 Buyers' behavior and corresponding beliefs in PM 75
3.25 Voluntary participation by treatment 76
3.26 Simulation Screen in BASE for truth-telling strategies 77

4.1 Payment dependence on type reports for two types 94
4.2 Ex post payoff profiles for truthful revelation 99

5.1 Comparisons between the auctions 115
5.2 The dependence of the bonus function on the conditional probability 118
5.3 The effect of the correlation on expected payoffs and profits 120
5.4 Expected profits for the auctioneer 125

List of Tables

2.1 Insurance and claiming predictions depending on risk and lying attributes . . . 17
2.2 Marginal effects of logit regression on cheating behavior . . . 26
2.3 Suggested model for insurance purchase prediction by stepwise command . . . 35

3.1 Equal share direct SCF $f^e(\theta_1, \theta_2)$ . . . 43
3.2 Equal share indirect mechanism Γ^e_{ind} . . . 43
3.3 Minimal subsidy SCF $f^*(\theta)$ . . . 46
3.4 The effect of reciprocity on the likelihood of truth-telling strategy in MSE . . . 56
3.5 Game induced by an direct mechanism of the SCF f^e . . . 66
3.6 Game induced by the indirect mechanism Γ^e_{ind} . . . 67
3.7 Game induced by an direct mechanism of the SCF f^* . . . 68
3.8 Price calculations . . . 82
3.9 Payment calculations . . . 82

4.1 Payoff tables depending on the valuation type profile θ . . . 96
4.2 Ex post utility function for positive correlation . . . 107
4.3 Ex post utility function for negative correlation . . . 108

5.1 Placement into the current literature . . . 113

Chapter 1

Introduction

In many markets, economic agents possess private payoff-relevant information. This asymmetric distribution of information, also often referred to as hidden information, affects the performance of the market mechanism. A general conclusion in the literature, is that markets with hidden information achieve an inferior outcome as compared to the same market where information is public. Classical examples, where private information leads to an inefficient allocation, are the quality of goods offered in second-hand markets (Akerlof, 1970), the personal skill in the labor market (Spence, 1973) or the risk traded in an insurance market (Rothschild and Stiglitz, 1976; Stiglitz, 1977). In these examples, the analysis focuses on the welfare consequences of the market failure, known as adverse selection. The general problem is that the informed market side has no incentive to truthfully reveal the possessed information, if a false claim increases the own material well-being. For instance, claiming to offer a high quality good, although it is actually of low quality, in a second-hand market would allow to ask for a higher price. If the uniformed market side, which is unable to learn the quality before the trade, anticipates such behavior, only low prices would be offered. This ensures that the uninformed side feels no regret from the transaction. But if there is only a low willingness to pay by the uniformed market side, high quality goods are not offered in the market anymore. Hence, only goods of low quality would be traded. Most of the theoretical literature in such an environment discusses (governmental) interventions that could moderate the consequences of adverse selection. Nevertheless, empirical evidence concerning the unraveling of (insurance) markets due to asymmetric information is mixed (Cohen and Siegelman, 2010). While several studies confirm adverse selection (for instance Puelz and Snow, 1994), other empirical evidence points in the direction of none (Chiappori and Salanie, 2000), or even advantageous selection (Fang et al., 2008). A general problem of field data to address these questions is that private payoff-relevant information is also not observable for researchers. The private information of interest needs to be measured with an observable variable. This thesis employs an empirical method that is able to overcome this fundamental problem - a controlled lab experiment. Before I discuss theoretical and empirical methods applied in this thesis, I introduce shortly the concrete environments considered in each Chapter and how they are connected to the examples above.

1.1 Environments

In the second Chapter of this thesis my coauthor and me analyze an information asymmetry in an insurance market. In this environment, every policy holder observes the size of her insured loss while the insurance company is unable to acquire this information (Shavell, 1979; Townsend, 1979). In theory the same consequences arrive as described above because it is in the self-interest of each policy holder to claim the highest possible monetary amount, independent of the actual materialized loss. Given that insurance companies anticipate this behavior, they should be unwilling to offer any policy for the risk. As a consequence, the insurance market would break down.

So far, I assume that only one side of the market has access to payoff-relevant information. I study a more general approach in Chapter 3 of this thesis: the bilateral trade environment where both market participants possess private payoff-relevant information (Chatterjee, 1982). An example for such an environment is a seller who knows the costs, while the buyer is informed about the valuation, but not vice versa. If private informations are uncorrelated[1], it is not possible to achieve efficient trade without an external subsidy or enforced participation (Myerson and Satterthwaite, 1983). An advantage is that the general insights from the bilateral trade environment are transferable to other problems, like the one of an efficient provision of a public good.

Another example that I investigate in Chapter 4 and 5 of this thesis has the objective to maximize the (expected) profit of a seller, who lacks information about the private valuations of potential buyers. If there is a single buyer, the literature shows that the posted price is in expectation among the best selling format a seller can apply. Here, I consider cases where at least two potential buyers are present. For these cases auction theory provides the most profitable format to sell the good. A famous result, the revenue equivalence theorem (Vickrey, 1961; Riley and Samuelson, 1981),

[1] Uncorrelated informations mean that the information one informed agent receives is independent of the information that another informed agent possesses.

shows theoretical evidence that there are no differences between different auction formats given uncorrelated valuations. The violation of the uncorrelated valuation assumption is the core interest in these Chapters. The main conclusion of environments where private valuations are correlated, such that their belief distributions are linear independent, is that the same expected profits are realizable as in the case that bidders' valuations are public information (Crémer and McLean, 1985, 1988). In the following, I will introduce the methods applied in the current thesis.

1.2 Methods

1.2.1 Game Theory and Mechanism Design

A general game theoretical approach to model these incomplete information environments are so-called **Bayesian games** (Harsanyi, 1967, 1968a,b). In its stage game form, the Bayesian game expands a non-cooperative complete information game[2] by private payoff-relevant information types for each player. In these games players only know the probability distribution of other players' private information and share a common (ex ante) belief. The corresponding Bayes Nash equilibrium is a modification of the Nash equilibrium (Nash, 1951) such that optimal strategies are constructed with respect to beliefs over information. A strategy is a mapping from each of the potential information types to an action. A Bayes Nash equilibrium requires that in expectation over private informations of other players, the strategies of all players are best responses with respect to each other.

In general, the field that applies game theory, and in particular Bayesian games, to study outcomes and interventions in environments with information asymmetry is known as **mechanism design** (Hurwicz, 1973). The analysis of a mechanism design problem is centered around the implementation of so called social choice function. For any possible realization of private information, a social choice function returns an allocation. What exactly an allocation is, depends on the environment. In general, it determines the distribution of existing resources among market participants, also called agents. For example, in an auction the allocation states the monetary amounts each bidder has to pay to the auctioneer and who of them receives the good. Studies in the mechanism design literature provide answers about which social choice

[2]A complete information game consists of a set of players, each with an associated set of actions and corresponding payoff functions (Nash, 1951). A Nash equilibrium is an action profile such that the actions that are best responses to each other. In other words, the actions chosen by each player yields the highest possible payoff given the actions of the others. In his fundamental work on game theory, Nash (1951) proved that in every complete information game a (mixed) Nash equilibrium exists.

functions are implementable. For the implementation they apply mechanisms which consist of a set of strategies for each agent and an outcome function that maps these strategies into an allocation. Intuitively, a social choice function is implementable, if there exists an equilibrium in the game induced by such a mechanism, where the outcome function coincides with the allocation. The modification that ensures the existence of an equilibrium is based on the idea of incentive compatibility (Hurwicz, 1972). The mechanism therefore induces a game where it is in the self-interest of each agent to behave in the interest of the society.

Most studies consider the implementation of social choice functions in one of two equilibrium concepts: The Bayes Nash equilibrium introduced earlier and the more restrictive equilibrium in Dominant Strategies[3]. Many optimal mechanisms applying Bayesian Implementation require that the designer has detailed knowledge about the beliefs about the private information among agents. In contrast, the later one is robust in the sense of Wilson (1987)'s doctrine towards heterogeneity in beliefs of agents. But the problem with such an implementation of a social choice function is stated in the Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975): the implementation in Dominant Strategies and arbitrary preferences is in general impossible. Exceptions of this impossibility are the commonly used quasi-linear preferences. In combination with pure private valuations, under the assumption agents' preferences are quasi-linear, Bergemann and Morris (2005) show an equivalence between Bayesian and Dominant Strategy implementation. Nevertheless, these preferences are centered around the self-interest hypothesis, i.e. selfish preferences. Criticizing the dependence of this assumption Bierbrauer and Netzer (2016) seek robustness towards heterogeneity in social preferences. In particular, they provide a condition that allows for robustness against the heterogeneity in social preferences.

1.2.2 Behavioral Economics

As mentioned before, the mechanism design literature works with the modification of Bayesian games which ensures that the self-interest of each agent is aligned with the interest of the whole society under the efficiency objective. On a wider range in economics, this self-interest hypothesis has been criticized by the behavioral economics literature (Güth et al., 1982; Kahneman et al., 1986). As an alternative, this literature introduces different utility functions representing preferences that are

[3]An equilibrium in Dominant Strategies requires that the strategy is optimal with respect to any possible strategies other agents play. In other words, independent of the strategy the other player plays, the equilibrium strategy remains a best response. This property ensures against strategic uncertainty of other agents.

empirically more plausible, at least on the individual level (Rabin, 1993; Fehr and Schmidt, 1999). Based on these utility functions the literature provides new equilibrium concepts that increase the possibility to predict and explain the outcome in games. In the last decade, many of these new results are expanded to increase the predictability also under information uncertainty, i.e. in Bayesian games.

Rabin (1993) introduces a fairness equilibrium within the psychological game theory literature (Geanakoplos et al., 1989), where players additionally to their self-interest also care about beliefs and the corresponding intentions of other players. The fairness equilibrium is defined for a two-person stage game. The proposed utility function incorporates the idea that if a player believes that the other player's action is kind to her, she is willing to give up material payoff in order to be also kind to the other player. But of course, also the opposite might be the case. If she believes that the other player is unkind to her, then she is willing to give up material payoff in order to be unkind to her. The equilibrium demands, first, that the actions are best responses with respect to the utility function, and second, that beliefs about the behavior and beliefs of the other player are correct. Another part of this literature focuses on the idea of outcome-based social preferences (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). Here, agents do not only care about the consequences of their behavior for their own payoff, but compare it with the effect on the other agents' payoffs. Hence, agents' behavior and the according equilibrium concepts are based on their relative payoffs. Saito (2013) expands the basic idea of Fehr and Schmidt (1999) to environments with a (non-strategic) background risk, as in Bayesian games, and distinguishes between the concepts of ex ante (procedural) and ex post (outcome) fairness. In general, these utility functions and equilibrium concepts are able to explain empirical findings, which are not plausible under the pure self-interest hypothesis.

More recently, behavioral economists also discuss the importance of social norms in economics, for example for the role in equilibrium selection (Young, 2015). A specific social norm of interest for information economics is the unwillingness of agents to lie. Recent studies report that people are unwilling to claim a false information to increase their monetary gain (see for instance, Abeler et al., 2016; Khalmetski and Sliwka, 2017). This contradicts the core idea of information economics and might explain the mixed empirical findings in insurance markets reported earlier. Notice that the finding is more relevant if the application is a direct mechanism, for instance in the case of insurance fraud (as in Chapter 2) than for the general mechanism design literature. There, in most cases, the revelation principle allows to focus on the truth-telling equilibrium in the direct mechanism. But for real world applications, the direct mechanism is solely an abstraction that represents the function of institutions, where the information revelation is rather considered as a strategy than a 'real' lie.

Most of the empirical findings, the behavioral economic literature is based on, are provided by studies in experimental economics. In the first two Chapters of the thesis I apply an experimental method for the empirical analysis. I continue therefore with a general introduction of experimental economics as the empirical method applied in the current thesis.

1.2.3 Experimental Economics

With his induced value theory, Smith (1976) introduces a theoretical foundation to apply incentivized controlled laboratory experiments to test economic theory. The basic idea is that preferences are induced due to the monetary compensation subjects receive for their participation in the experiment. Hence, the designer of the experiment is also able to implement private payoff-relevant information. In general, with controlled laboratory experiments, one has full control of the information provided to each experimental subject and the (Bayesian) game they play, while the behavior (and beliefs) is not predetermined. This controllability provides a tool that enables research for the investigation of insights provided by the mechanism design literature.

Smith (1962) himself was among the first who used this method to verify the market prediction of the neoclassical theory in an isolated environment. He provides evidence that in competitive markets, despite the presence of hidden information on both market sides, prices converge to the competitive equilibrium price over time. In this thesis, I apply the experimental method for the same motivation: the observation of individual behavior of market sides who possess private payoff-relevant information and the consequential outcome in the market. In other words, I conduct an empirical test in an isolated environment which allows to identify the effect of an institutional modification. This differs from the second possible application of laboratory experiments: to identify channels that determine human behavior (see the psychological literature).

Experiments that identify the underlying factors of human behavior lead to the development of social preference models in economics in the first place. Experimental economists design games and experiments that allow to disentangle explanatory factors of behavior. A famous example is the dictator game (Kahneman et al., 1986) which modifies the ultimatum game (Güth et al., 1982) such that the receiver has no longer a choice to accept or decline the proposer's offer. In contrast to the original game, only social preferences can be accountable for the observed deviation from the subgame perfect equilibrium (Selten, 1975) based on selfish preference. In general, the extensive literature of experimental evidence using dictator games provides evidence for a causal link between outcome-based social preferences and decision making (Engel, 2011). Falk et al. (2003) report experimental evidence that there exists an isolated effect of reciprocal motives. Concerning lying aversion, Gneezy (2005) finds that experimental subjects are unwilling to lie in a strategic situation if the material payoff from it is small. In contrast, Sutter (2009) shows that subjects use 'sophisticated' deception in sender-receiver experiments. Hence, senders report truthfully under the assumption that the receiver does not follow the report. In a non-strategic pure decision making task, Fischbacher and Föllmi-Heusi (2013) provide evidence in a double-blind experiment that subjects are indeed unwilling to lie.

1.2.4 Mechanism Design and Social Preferences

In the sense of Smith (1962), the experimental method has been applied for the verification of results from the mechanism design literature. This literature provides insights which specific mechanisms achieve their objective. If a mechanism fails to implement it, experiments can be useful to provide reasons for the failure. Chen (2008) provides a survey of the experimental literature concerning the optimal provision of public goods. She concludes that in pivotal mechanisms misrevelation is prevalent (Attiyeh et al., 2000; Kawagoe and Mori, 2001), the results with refinements of the Nash equilibrium are ambiguous and that mechanisms that induce supermodular games[4] converge to the Nash equilibrium. In the domain of optimal auction Kagel and Levin (2016) provide a broad overview of the extensive literature. One main motivation for the literature was to explain why the revenue equivalence theorem does not hold empirically (Coppinger et al., 1980; Cox et al., 1982; Kagel et al., 1987). Experimental work is helpful to identify the underlying reasoning of the observed overbidding in comparison to the predicted (Nash) equilibrium in many auction formats (first price and second price auction). More recent work focuses on theoretical and experimental insight for the application of auctions in public and private institutions, for instance for spectrum auctions. With respect to the influence of social preferences on the performance of mechanisms, there is experimental

[4]"Supermodular games are games in which each player's marginal utility of increasing her strategy rises with increases in her rival's strategies, so that (roughly) the player's strategies are "strategic complements". Supermodular games have very robust stability properties, in the sense that a large class of interesting learning dynamic converges to the set bounded by the larges and the smallest Nash equilibrium strategy profiles." - Chen (2008)

evidence in auctions by Bartling and Netzer (2016) and for income taxation as well as bilateral trade by Bierbrauer et al. (2017).

The implication of social preferences for markets with information asymmetries depends on the environment. Results from the theoretical literature range from efficiency gains due to pro-social behavior for the optimal provision of public goods and private good trade (Kucuksenel, 2012) to destructive spite in auctions (Morgan et al., 2003). Bierbrauer and Netzer (2016) show how reciprocal preferences based on Rabin (1993) help to overcome the Impossibility Theorem by Myerson and Satterthwaite (1983) and describe circumstances where the revelation principle no longer holds.

In general, this thesis provides experimental and theoretical insights how social preferences of privately informed agents affect the market outcome. In the next three Chapters I show that using social preferences to ensure the implementability of social choice functions is difficult due to their heterogeneous distribution within the population. The last (and partially the third) Chapter contribute to the effect of social preference robustness conditions.

1.3 Chapter summaries

In the following I introduce four Chapters of the thesis. I state the question asked within the study and provide an overview of the results. Each study applies at least one of the two introduced methods and studies an information economics problem. In detail, I investigate the effect of at least one kind of social preferences in environments where agents possess private payoff-relevant information.

Chapter 2: *Compulsory versus Voluntary Insurance: How Leaving Choice Affects Fraudulent Behavior (based on joint work with Franziska Tausch)*

In this Chapter my coauthor and me investigate whether the circumstances under which an insurance contract is concluded affects ex post moral hazard. In a controlled laboratory experiment we compare false loss reporting behavior by policy holders under compulsory insurance to a setting in which individuals can freely choose their insurance coverage. The standard selfish preference predicts that each policy holder reports the highest possible claim because it is first order stochastic dominant and there are no differences between the different contract types. In contrast, we find that policy holders deviate from this standard prediction. Indeed, cheating is significantly higher under voluntary insurance and that this effect is driven by the selection of fraudulent individuals into the insurance contract. Our results suggest that compulsory insurance is not only an effective measure to avoid adverse selection of individuals that are particularly likely to claim actual losses, but also the selection of those that are likely to claim false or exaggerated losses.

Chapter 3: *The Role of Intention in Bilateral Trade Environments: An Experiment*

In a controlled laboratory experiment, I study the role of intentions among privately informed market participants in a bilateral trade environment. Contrary to theoretical insights by Bierbrauer and Netzer (2016), I do not find empirical support for their counterexample to the revelation principle. The authors show that the implementation of a social choice function equally shares the gains of trade the authors equal share of trades social choice function. The modification increases, as predicted, the perceived kindness of the truth-telling strategy, but I conclude that the unsuccessful implementation is due to the decreasing trust towards sellers to behave kindly. Although there is significantly less truth-telling in this indirect mechanism compared to the direct one, I find no differences in the frequency of efficient trade between the two mechanisms. The reasoning here is that in the indirect mechanism multiple equilibria lead to the efficient trade. I also conclude that there are no differences with respect to subjective well-being between the mechanisms.

Chapter 4: *The Dependence of Crémer-McLean Auctions on Selfish Preferences*

In the fourth Chapter of the thesis I study the effect of outcome-based social preferences on auction design in correlated environments. I consider two bidders with two possible valuation types who bid for a single unit object. I show that in general the auction by Crémer and McLean (1985) is not robust against outcome-based social preferences. In the standard case of an indivisible good selfish preferences are not only sufficient but also necessary for the existence of a truth-telling ex post equilibrium. The binding incentive-compatibility for both valuation types permits the possibility to affect the ex post payoff of the other bidder without consequences for the own ex post payoff. I consider two less restrictive cases: the ex post implementation of a divisible good and Bayesian implementation. For these cases I conclude that uncertainty over the distribution of outcome-based social preferences increases the volatility of the expected profit for the auctioneer.

Chapter 5: *Social Robust Auctions: The case of correlated valuations*

The last Chapter of the thesis investigates the effect of the externality-freeness condition on the optimal design of auctions under the assumption that valuation types of bidders are correlated. Again, I consider two bidders with two possible valuation types who bid for a single unit object. Bierbrauer and Netzer (2016) introduce the externality-freeness condition to ensure robustness with respect to an unknown heterogeneity of social preferences among bidders. I consider ex post and Bayesian incentive compatibility and relate the results to insights in the literature. In general, I show that the first best implementation is no longer possible under the externality-freeness constraint. For the case of Bayesian incentive compatibility, I find a- continuous effect of the intensity of correlation on the auctioneer's expected profit. Under ex post incentive compatibility there are no differences for the optimal auction design given correlated and uncorrelated valuation types.

Chapter 2

Compulsory versus Voluntary Insurance: How Leaving Choice Affects Fraudulent Behavior

2.1 Introduction

Policy makers naturally attempt to improve the allocation of resources with the aim to increase overall well-being. In the insurance context, two types of inefficiencies may occur that relate to a lack of insurance demand among particular groups of individuals. First, mostly high risk individuals may choose to insure (e.g., Tausch et al., 2014, Cutler and Zeckhauser, 1998) which threatens the sustainability of effective risk sharing arrangements through increased insurance prices. Second, individuals may neglect to insure completely or under-insure despite failing to reach their optimal insurance coverage (e.g., health: Lavarreda et al., 2011, Blewett et al., 2006, catastrophic risk: Kunreuther, 1984, automobiles: Findling and Germano, 1988).

A classical example for an intervention by the state that counteracts those issues is to introduce the legal obligation to purchase insurance (Rothschild and Stiglitz, 1976; Wilson, 1977). That way equality in insurance access for all risk types can be secured and under-insurance can be avoided. While such a paternalistic intervention may run into opposition as it deprives individuals of their freedom of choice, a more libertarian approach is increasingly applied which is the attempt to nudge individuals into purchasing (more) insurance. For example, insurance may be included into a purchase unless the customer explicitly declines it, or default options are specified or preselected whose choice presumably entails less effort for the customer.

The aim of this study is to investigate whether insurance favoring interventions entail hidden costs in the form of increased moral hazard among the insured.

The conclusion of an insurance contract, irrespective of whether it is voluntarily or compulsory, may implicate unproductive behavior: policy holders behave carelessly, don't invest in risk prevention (ex ante moral hazard; see Hölmstrom, 1979; Shavell, 1979), make claims to the insurance company that are higher than their actual loss or they do not take the least costly measure to eliminate an actual damage (ex post moral hazard; see Townsend, 1979; Gale and Hellwig, 1985; Lacker and Weinberg, 1989). We analyze how the obligation to be insured and being nudged into an insurance contract affects ex post moral hazard. In particular, we investigate how leaving individuals the choice whether to insure or not affects claim build-up and fictitious claiming, i.e. the extent to which policy holders make exaggerated claims after risk realization.

Applying the experimental methodology in this study allows us to track fraud on an individual level. We observe the actual size of the individual loss and can match it with the loss amount that is claimed. Furthermore, we can keep the context neutral and compare settings that differ only with respect to the insurance process. This allows us to clearly identify how the circumstances under which a contract is concluded affect (dis)honesty.

Our results reveal that insured individuals cheat significantly less under compulsory as compared to voluntary insurance and that this effect is driven by the selection of fraudulent individuals into the insurance contract. We conclude that compulsory insurance is great because it grants equal access for all risk types to insurance contracts and additionally, the risk sharing arrangement is more stable due to the lower fraction of fraudulent individuals as compared to a voluntary setting.

This research project relates to the behavioral economics literature on lying aversion which is based on the idea that individuals like to perceive themselves as honest and like to be perceived as such by others (e.g., Abeler et al., 2016; Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013). A widespread finding is that a significant fraction of individuals are unwilling to lie for a monetary benefit. Whether the reluctance to behave dishonestly extends to an insurance context and is influenced by the circumstances under which an insurance contract is concluded is an empirical question.

Another factor that has been investigated to influence fraudulent behavior is the incentive structure of contracts. In an experimental study Lammers and Schiller (2010) find more fraudulent behavior in a deductible setting than in a full insurance condition. In a dynamic environment Gabaldón et al. (2014) report no difference in cheating behavior between a bonus-malus contract and under a classical audit system while von Bieberstein and Schiller (2017) find a substantial increase of insurance fraud in a deductible contract in comparison to one using a bonus-malus system. Our study furthermore adds to the literature on side effects of nudging. Handel (2013) shows that nudging in the form of information provision leads to an unraveling of the insurance market in that less people purchase comprehensive coverage. Damgaard and Gravert (2018) show that nudging in the form of sending reminders leads people to drop out of the mailing list.

2.2 Theoretical Framework

We introduce a theoretical framework that is the base from which we derive hypotheses for the extent of insurance claim exaggeration. We contrast cheating across three treatments in which the conclusion of the insurance contract is purely voluntary (Vol), voluntary but individuals are nudged towards the insurance policy (Nudge) or compulsory (Comp).

We consider an insurance company that offers the following insurance policy (p, f) to an individual: *"In exchange to a price p paid in $t = 1$, the individual has the right to claim compensation f of a loss l with $l = f$ in $t = 2$."* The individual i is initially endowed with w_0, faces a risky loss represented by random variable l with support on $L = \{0, \ldots, \bar{l}\}$ and probability mass function $q : L \to [0, 1]$. Once the insurance policy is offered on the market the individual decides in $t = 1$ whether to obtain an insurance policy ($a \in A$) which allows to claim compensation $f \in L$ after privately observing the outcome of the loss $l \in L$.[1] We manipulate the choice set A in our analysis such that

$$A = \begin{cases} A_V = \{0, 1\} & \text{if voluntary,} \\ A_N = \{\mathbf{1}, 0\} & \text{if nudged,} \\ A_C = \{1\} & \text{if compulsory.} \end{cases}$$

The ex post wealth level of the individual composes of the initial wealth w_0, loss l, price p and claim f if an individual is insured,

$$w = \begin{cases} w_0 - l - p + f & \text{if a} = 1, \\ w_0 - l & \text{if a} = 0. \end{cases}$$

[1] A crucial assumption we make throughout the analysis is that the verification of the loss is never possible such that we exclude possible verification methods discussed in the insurance fraud literature (Townsend, 1979; Gollier, 1987). Thus deviation from predictions based on standard preferences are not explainable by effects of monetary sanctions in the case of discovery.

We assume a (strictly) concave utility function over wealth $u(w)$ which accounts for the degree of risk aversion and adapt the approach by Khalmetski and Sliwka (2017) to introduce lying aversion into the utility function. Suppose first an individual obtains the insurance policy. Then preferences of the individual in $t = 2$ are represented by

$$U_{i,2}(w, f, l, \gamma_i, \eta_i) = u_i(w) - \gamma_i \; 1_{\{f>l\}} - \eta_i \; Pr[l \neq f|f]. \tag{2.1}$$

The first term represents the utility from wealth. In the second part $-\gamma_i \; 1_{\{f>l\}}$ illustrates the lying cost from misreporting that results from hurting the self-image. This psychological cost occurs when the claim exceeds the actual loss.[2] Furthermore, $-\eta_i \; Pr[l \neq f|f]$ represents a dis-utility from hurting the social-image through not being perceived as honest by other people. This in turn depends on the likelihood of a false report, i.e. $l \neq f$, conditional on report f. We assume that the probability of misreporting increases with the size of the claim (see Abeler et al., 2016, for empirical support) .

The utility function in $t = 1$ is the expectation over potential ex post utilities

$$U_{i,1} = \sum_{l \in L} q(l) \; U_{i,2}(w, f, l, \gamma_i, \eta_i). \tag{2.2}$$

If the individual is not insured there is no opportunity to misreport and preferences are simply represented by $U_{i,2} = u_i(w)$. The profit of the insurance company is $\Pi = p - f$ if the individual insures and zero otherwise.

2.2.1 Standard Preferences: $\gamma_i = 0$ & $\eta_i = 0$

Only if an individual i is insured ($a = 1$) a claim with the insurer can be made. In that case the utility in $t = 2$ after observing the actual loss is

$$U_{i,2}(w, f, l) = u(w_0 - l - p + \bar{l}) \tag{2.3}$$

since

$$\bar{l} \in \arg\max_{f \in L} \;\; u(w_0 - l - p + f). \tag{2.4}$$

An insured individual always reports the highest possible loss $\bar{l}$, irrespective of circumstances under which the contract was concluded as long as utility is increasing

[2] We exclude psychological costs that might occur through reporting less than the actual loss. We do not expect such behavior as there is no trade-off between monetary improvement and misreporting and importantly, we do not observe such behavior in our experiment.

in wealth. If the insurance choice $a \in A$ is voluntary or nudged, the loss reporting decision is preceded by the choice of the insurance policy in $t = 1$. Individual i purchases the insurance if and only if

$$\sum_{l \in L} q(l)\; u(w - l - p + \bar{l}) \geq \sum_{l \in L} q(l)\; u(w_0 - l)$$
$$\Leftrightarrow \quad \bar{l} \geq p. \tag{2.5}$$

We conclude that under the assumption of standard preferences individuals always purchase the insurance if given a choice and claim the highest possible amount, as long as this amount is larger than the price for the insurance.[3] The contractual circumstances are predicted to not affect cheating behavior.

2.2.2 Lying Aversion: $\gamma_i > 0$ & $\eta_i > 0$

Recent experimental evidence, however, shows that people forgo monetary payoffs to avoid lying (Abeler et al., 2016; Mazar et al., 2008; Fischbacher and Föllmi-Heusi, 2013). Based on the experimental data, several recent studies conclude that the reason for lying costs is a combination of a preference for being honest and a preference to appear honest (Abeler et al., 2016; Dufwenberg and Dufwenberg, 2018; Khalmetski and Sliwka, 2017). Following their conclusion, as in Khalmetski and Sliwka (2017) we assume an individual specific dis-utility es depicted in equation (2.1) with $\gamma_i, \eta_i \geq 0$. Consider first the case where the individual is insured. The optimal claim in $t = 2$ is

$$f^* \in \arg\max_{f \in L} u(w_0 - l - p + f) - \gamma_i\; 1_{\{f>l\}} - \eta_i\; Pr[l \neq f | f]. \tag{2.6}$$

Figure 2.1 represents different degrees of lying aversion. The optimal claim lies within the range $[l, \bar{l}]$.[4]

If the insurance choice $a \in A$ is voluntary or nudged, individual i makes the insurance policy decision in $t = 1$. An individual that anticipates his behavior in

[3] Anticipating this behavior no insurance would be offered in the first place (see Shavell, 1979).

[4] Given a very high η also optimal claims lower than l are possible. We do not observe such claim behavior in the experimental data which coincides with other studies as Gneezy et al. (2018) who find that there is almost no underreporting (1 out of 602 observations).

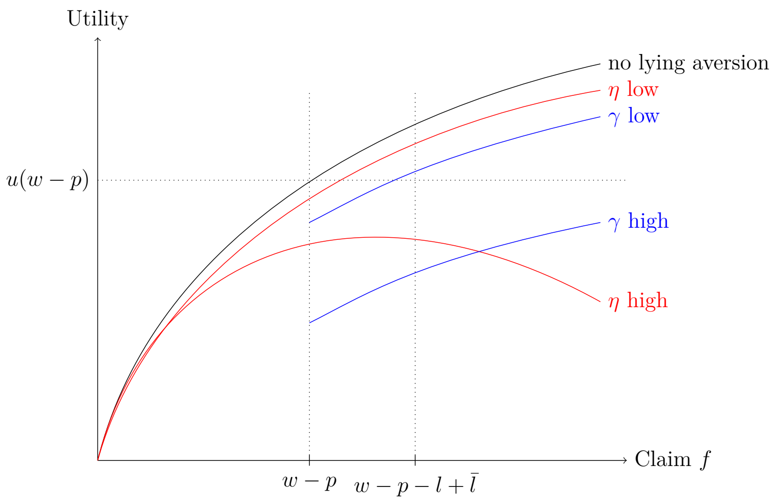


Figure 2.1: Ex post utility with lying aversion

$t = 2$ purchases the insurance policy if and only if

$$\sum_{l \in L} q(l)\ u(w_0 - l - p + f) - \gamma_i\ 1_{\{f>l\}} - \eta_i\ Pr[l \neq f|f] \geq \sum_{l \in L} q(l)\ u(w_0 - l)$$
$$\Leftrightarrow \sum_{l \in L} q(l)\ [u(w_0 - l - p + f) - \ u(w_0 - l)] \geq \gamma_i\ 1_{\{f>l\}} + \eta_i\ Pr[l \neq f|f]. \qquad (2.7)$$

As equation (2.7) illustrates the individual's decision to purchase the insurance depends on risk and lying preferences. An illustrating summary of the predictions depending on those two factors is depicted in Table 2.1. First, suppose that γ_i and η_i are defined such that the solution of equation (2.6) is $f = \bar{l}$, i.e. that and individual will cheat to the full extend (left column). In that case the analysis from the standard preference case repeats and individuals always purchase the insurance if given a choice and claim the highest possible amount, as long as this amount is larger than the price for the insurance. Next we consider cases where the optimal claim is lower than $\bar{l}$. For illustration we pick the truthful report $f^* = l$ (right column), but the argument generalizes for every $f^* \in (l, \bar{l})$. In this case the prediction depends on an individual's risk attitudes, i.e. the concavity of $u(\cdot)$. The more risk averse a highly lying averse individual, the more likely he is to insure.

Predictions	Low Lying Aversion	High Lying Aversion
Low Risk Aversion	Comp and Vol: Insurance (a=1) & $f^* = \bar{l}$	Vol: No Insurance (a=0) Comp: Insurance (a=1) & $f^* = l$
High Risk Aversion	Comp and Vol: Insurance (a=1) & $f^* = \bar{l}$	Comp and Vol: Insurance (a=1) & $f^* = l$

Table 2.1: Insurance and claiming predictions depending on risk and lying attributes

In case of a compulsory insurance all types of individuals are required to purchase the insurance. Hence, while individuals that intend to cheat will choose the insurance in the voluntary setting, in the compulsory setting both cheaters and non-cheaters are present. Similarly, some non-cheaters may be pushed into the policy through the nudge intervention when given the choice about the insurance purchase. Consequently, cheating should be highest in the Vol treatment, followed by the Nudge treatment and the Comp treatment.

2.2.3 Factors Influencing γ_i & η_i

Next to selection effects, we further consider direct effects of the contractual circumstances on cheating behavior. In particular, we discuss two potential mechanisms that might influence the degree of lying aversion, i.e. the size of γ_i and η_i.

Self-Serving Justification

When faced with the decision whether to behave dishonestly or honestly, people tend to interpret situations in a way that allows them to reap the benefits from dishonesty while only incurring low psychological costs through lying aversion (e.g., Shalvi et al., 2015). For example, if people perceive that an unethical act would reduce some kind of unfairness, they are more willing to misbehave (e.g., Shalvi et al., 2015, Fukukawa, 2002). In our context individuals may justify their dishonesty by pointing at the insurance premium that is higher than actuarially fair, i.e. it exceeds the expected loss (Tennyson, 1997, see Köneke et al., 2015, for further references). Similarly, individuals may engage in a form of moral licensing and use their 'good deed' of behaving cautiously by purchasing the insurance, to justify the 'bad deed' of making an exaggerated claim. We expect that the more conscious the insurance decision was made, the more likely people may be to justify their misbehavior. This mechanism would predict that cheating is highest in the Vol treatment, followed by the Nudge treatment, and lowest in the Comp treatment.

Control Aversion

Empirical evidence suggests that people are averse to being restricted in their choice set, and may thus negatively reciprocate to the entity that restricts them (e.g., Falk and Kosfeld, 2006). Individuals that are forced to insure could dislike the lack of an opt-out option and may retaliate with an exaggerated or faked insurance claim. The elimination of an element of the choice set in the Comp treatment is a stronger interference than the priming of an element in the Nudge treatment. The control aversion reasoning would thus predict that the cheating rate is highest in Comp, followed by the Nudge treatment and lowest in the Vol treatment.

2.3 Experimental Design

The experiment consists of two parts in which subjects' decisions are incentivized, followed by a questionnaire. The first part elicits individual risk preferences, while the second part contains an insurance experiment that reflects the theoretical framework we just introduced. At the beginning of the experiment subjects are provided with instructions for both parts. In order to insure the understanding of the instructions, they are required to answer a set of control questions. Only if all subjects had correctly answered all questions, we would start with the experiment.

2.3.1 Risk Elicitation

We implement a variant of the "Bomb Risk Elicitation Task" introduced in Crosetto and Filippin (2013). At the beginning subjects are required to work for the endowment that is subsequently used for the Bomb task. We employ a modified version of the real-effort task introduced in Benndorf et al. (2014). Subjects are asked to encrypt three combinations of three letters into numbers. Each letter has to be assigned a three digit number that can be read off a table on the same screen. For the three correctly encrypted letter combinations they earn 3 Euros in the form of 100 (virtual) boxes; each box is worth 3 Eurocents. One of the boxes contains a bomb. The computer would throw away the boxes one after the other and the subjects are asked to decide when to stop the computer from throwing away those boxes. If the box with the bomb was among those boxes that were thrown away, subjects could keep all the boxes which they hadn't thrown away. If, however, the box with the bomb was not among those that were thrown away, the bomb would explode and destroy all boxes such that subjects get zero earnings for this part of the experiment. Thus, with each box that is thrown away, the probability to receive zero is lowered by 1%, but at the same time the possible earnings are lowered by 3 Eurocents. The more risk averse, the later subjects would stop the computer from throwing away the boxes. Subjects do not receive information about their earnings from the risk elicitation task until the end of the experiment.

2.3.2 Main Experiment

At the beginning of the second part each subject takes part in the same effort-task as employed in the first part of the experiment. This time subjects needed to encrypt eleven combinations of three letters. They are paid 11 Euros for their work and are informed in advance that their income is exposed to the risk $\tilde{r}_x = (-6, \frac{1}{3}, -3, \frac{1}{3}; 0)$. After completion of the task subjects can decide whether they want to purchase a full coverage insurance for a price of 4 Euro. This corresponds to the fair price of 3 Euro plus a 33% mark up of 1 Euro. The parameters are chosen with the aim to have at least 50% of the subjects purchase the insurance.[5] We implement three treatments that vary with respect to the design of the insurance situation: Vol, Comp and Nudge. In Vol subjects can choose whether to purchase the insurance. In Comp subjects need to purchase the insurance and do not have a choice. In Nudge subjects can choose, but they are pushed into the direction of purchasing the insurance through a default intervention. We apply a between subject design such that each subjects only takes part in one of the treatments.

In the Vol treatment, subjects are asked to choose between two sealed envelopes: the no insurance envelope and the insurance envelope. Each envelope contains three matchboxes that are also sealed. They contain the incomes corresponding to the different risk realizations. While the no insurance envelope contains three matchboxes with either 11 (no loss), 8 (3 Euro loss) or 5 Euros (6 Euro loss), in the insurance envelope the 4 Euros insurance premium is additionally deducted, resulting in incomes of 7, 4 and 1 Euros. Subjects are informed that the insurance would allow them to receive a refund of their actual loss from the insurance. In particular, they can make a claim $f_i \in (0, 3, 6)$ to the insurance company, by simply indicating what loss they occurred. Subjects have to indicate their decision between the two envelopes on the computer screen. An experimenter then comes to a subject's seat, checks the indicated insurance decision and hands out the according envelope.

In the Comp treatment, subjects can not choose whether they want to purchase

[5]The parameters were pretested in a pilot study in which 20 subjects were provided a short description of the Vol treatment and asked to indicate (a) whether they want to purchase the insurance and (b) which insurance claim they want to make conditional on all possible actual loss outcomes. Incentives were down-scaled as compared to the main experiment such that subjects could earn a maximum of 5.50 €. 55% of the subjects decide to purchase the insurance. None of the subjects reports a loss that is smaller then the actual loss. In the condition of a 6 € loss all subjects claim 6 €. In the condition of a 3 € loss 73% claim 6 € instead of their actual loss. In the condition of a zero loss 36% claim 6 € and 18% claim 3 €.

insurance or not. In order to keep constant between the treatments that subjects are in contact with the experimenter when the envelope is handed over, also in Comp the insurance envelope is distributed at the same point in time in the experiment as in the Vol treatment.

In the Nudge treatment, subjects are asked to choose whether they want to purchase the insurance or not, but they face an insurance favoring default, i.e. they start out with the insurance envelope in their cubicle and can decide whether they would like to reject the insurance by exchanging the envelope for the no insurance envelope. Also, on the computer screen the option in favor of the insurance is pre-ticked. Subjects are informed that an experimenter will drop by each cubicle, check on the insurance decision and exchange the envelope if this is desired.

In all treatments, subjects are then asked to open one of the three matchboxes privately in the envelope in their cubicle and to collect the money. The draw reflects the 1/3 chance of either incurring a high loss, a low loss or no loss.[6] Money in the matchboxes is divided into coins and notes such that subjects could not infer the content of a box. The earnings in the different matchboxes are denominated as follows: 5 Euro note + 2x2 Euro coin (11 Euro), 5 Euro note + 3x1 Euro coin (8 Euro), 2x2 Euro coin and 1 Euro coin (5 Euro), 5 Euro note + 2x1 Euro coin (7 Euro), 2x2 Euro coin (4 Euro), 2x0.5 Euro coin (1 Euro).[7] Individuals that drew the insurance envelope then decide whether and which claim $f_i \in (0, 3, 6)$ they want to make to the insurance. They indicate their decision on the computer.

At the end of the experiment, subjects are paid out their indicated insurance claim, the show-up fee and the earnings from the risk elicitation task.

2.3.3 Questionnaire

Subsequent to the main experiment, subjects are asked to answer a questionnaire that includes three sets of questions concerning (1) Demographics (2) Lying and norm-violating behavior (3) (Soft) paternalistic preferences and self-determination in decision making. The latter includes variables that have been suggested to be correlated with individuals' risk preferences in previous studies.

[6]Pooling the observations from all treatments we find that the distribution of actual losses is not different from a uniform distribution (Pearson chi^2, p=0.293). We provide an overview in Appendix 2.C.2.

[7]Note, that the lowest possible income is 11-6-4=1 Euro (purchase the insurance, incur the maximum loss and do not claim anything) and the highest is 11-0-4+15=13 (purchase the insurance, do not incur a loss and claim the maximum loss).

2.3.4 Experimental Procedure

The experiment was conducted in December 2016 at the *Kölner Laboratorium für Wirtschaftsforschung* at the University of Cologne. Subjects were recruited on-line with hroot (Bock et al., 2014). The software implementation was done with z-Tree (Fischbacher, 2007). A typical session lasted approximately 43 minutes and the average earnings were 10.26 €, including a 4 € show-up fee. In total 130 subjects participated in five experimental sessions (54 in Vol, 24 in Comp and 52 in Nudge). In order to ensure privacy, subjects could open their drawn matchbox in the cubicle and immediately pocket the money. Through this procedure it was impossible for the experimenter to know during the pay-out whether a subject cheated or not. While thus privacy is ensured during the experiment, we required subjects to leave the two unopened matchboxes in their cubicle in order to verify their actual losses after the experiment (see Friesen and Gangadharan, 2012, 2013, for applications of that procedure). Two observations had to be dropped from our sample in the Comp treatment, as the subjects left the money from their drawn matchbox in their cubicle. The experimental instructions and the questionnaire translated from German can be found in the Appendix.

2.4 Results

2.4.1 Insurance Purchase

While in the *Comp* treatment all subjects are insured by design, in the *Vol* and *Nudge* treatment we observe that 63% and 67% of the subjects respectively decide to purchase the insurance. Obviously, the nudging intervention was not effective in increasing insurance take-up as compared to its voluntary benchmark. We therefore focus our analysis mainly on the comparison between the Vol and the Comp treatment.

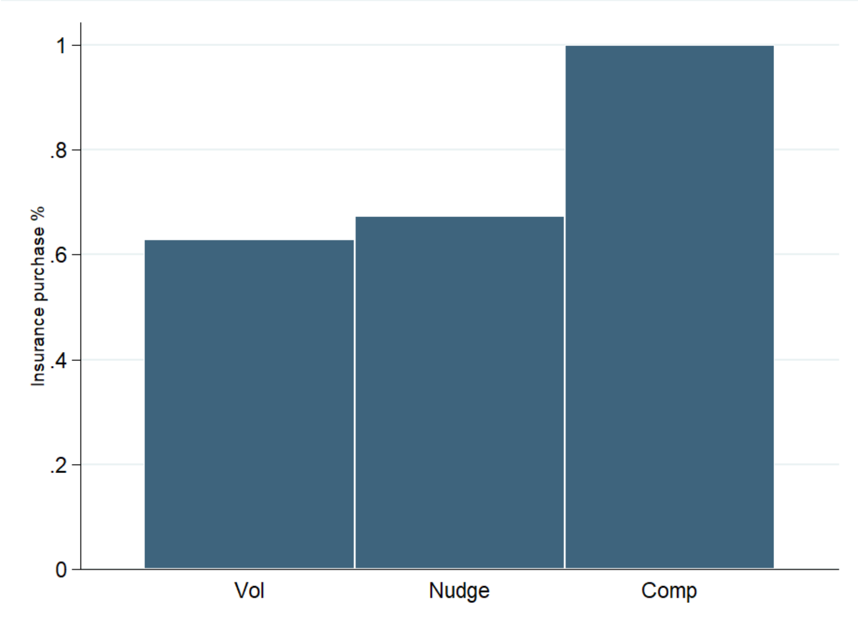


Figure 2.2: Cheating rate by treatment

2.4.2 Cheating

In order to investigate ex post moral hazard we compare subject's actual losses with the claims they indicated after the risk realization. Cheating is defined as reporting a loss that is larger than the actual loss. Note, that subjects with an actual loss of 6 € do not have the scope to exaggerate their claim and can therefore not be considered in the following analysis.

We first consider cheating at the extensive margin. Figure 2.2 depicts the percentage of cheaters across treatments. We observe that subjects cheat less in the Comp treatment as compared to the Vol treatment, both if the actual loss is zero or 3 €.

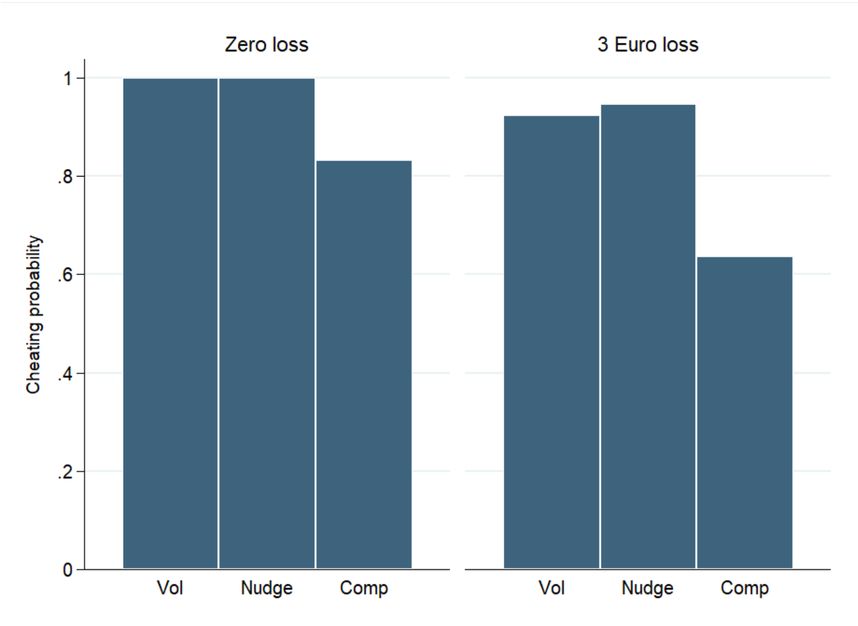


Figure 2.3: Cheating rate by treatment

Pooling both categories, a Fisher-exact test yields that differences between the two treatments are significantly different (p=0.014).[8] Splitting the analysis yields, that in both conditions differences are insignificant (p≥0.142), which however is likely due to the low number of observations (Comp: 12 and 13, Vol: 6 and 11). The logit regression in Table 2.2 column 1 that controls for individuals' actual losses confirms that cheating is significantly lower when insurance purchase is compulsory as compared to voluntary. In particular, subjects are 23 percentage points more likely to cheat in Vol as compared to Comp.

Result 2.1. *The percentage of individuals who exaggerate their insurance claims is significantly higher under a voluntary as compared to a compulsory insurance setting.*

Considering cheating at the intensive margin we find that all of the subjects in Comp cheat fully, i.e. they report a loss of 6 €, irrespective of their actual loss. Similarly, only one of the subjects in Vol that does not incur any loss cheats partially and claims 3 € only. None of the subjects claims a loss that is lower than the actual loss.

Since the nudging intervention did not trigger higher insurance-take-up it is not surprising that the cheating rate in Nudge is not different from that in Vol (Fisher exact test, p=1.00) and significantly higher as compared to Comp (Fisher exact test, p=0.018). Among the subjects who cheat, all but one subject who did not incur any loss claim the highest loss of 6 €.

[8]All reported tests are two-sided.

As described in Section 2.2 two channels may explain the higher cheating rates in the voluntary as compared to the compulsory insurance setting: a higher self-justification to cheat due to the self-determined purchase of the insurance or the selection of a particular type of subjects into the insurance contract.[9] In order to disentangle the two channels we use the risk elicitation results, the questionnaire data on subjects' demographics and their information on lying and norm-violating behavior as independent variable to find a linear logit model that best predicts the likelihood of voluntary insurance purchase for subjects in the Vol treatment. A stepwise backward-selection estimation that removes terms with $p \geq 0.1$ suggests a model that includes a participant's gender, age, the belief about the percentage of cheaters among other participants ("Indicate your belief about how many percent of the other participants that are insured claimed an amount hight than their actual loss"), self-stated previous cheating as a seller ("Did you ever lie to sell something?") and self-stated previous cheating as an applicant ("Did you ever lie in an application for work, a membership, school, university or foundation"). The model has a Pseudo R^2 of 0.43.[10] We then use the estimated coefficients of this model to predict the probability that participants in Comp would have bought the insurance if they had had the opportunity to choose freely.

[9] Pairwise correlations between the cheating behavior under compulsory insurance and participants' attitudes towards (soft) paternalism confirm that the control aversion channel does not play a role in our setting. None of the correlations is statistically significant.

[10] The model is further supported when we apply the leaps-and-bounds algorithm by Furnival and Wilson (1974) (programmed in Stata by Lindsey and Sheather, 2010) which performs the variable selection assuming a linear regression. Both, the Bayesian information criterion (BIC) and Akaike's corrected information criterion (AICC) confirm the model suggested by the stepwise approach. Akaike's information criterion (AIC) suggests to further include the number of siblings and the education level. The following results are robust to including these additional independent variables in the analysis.

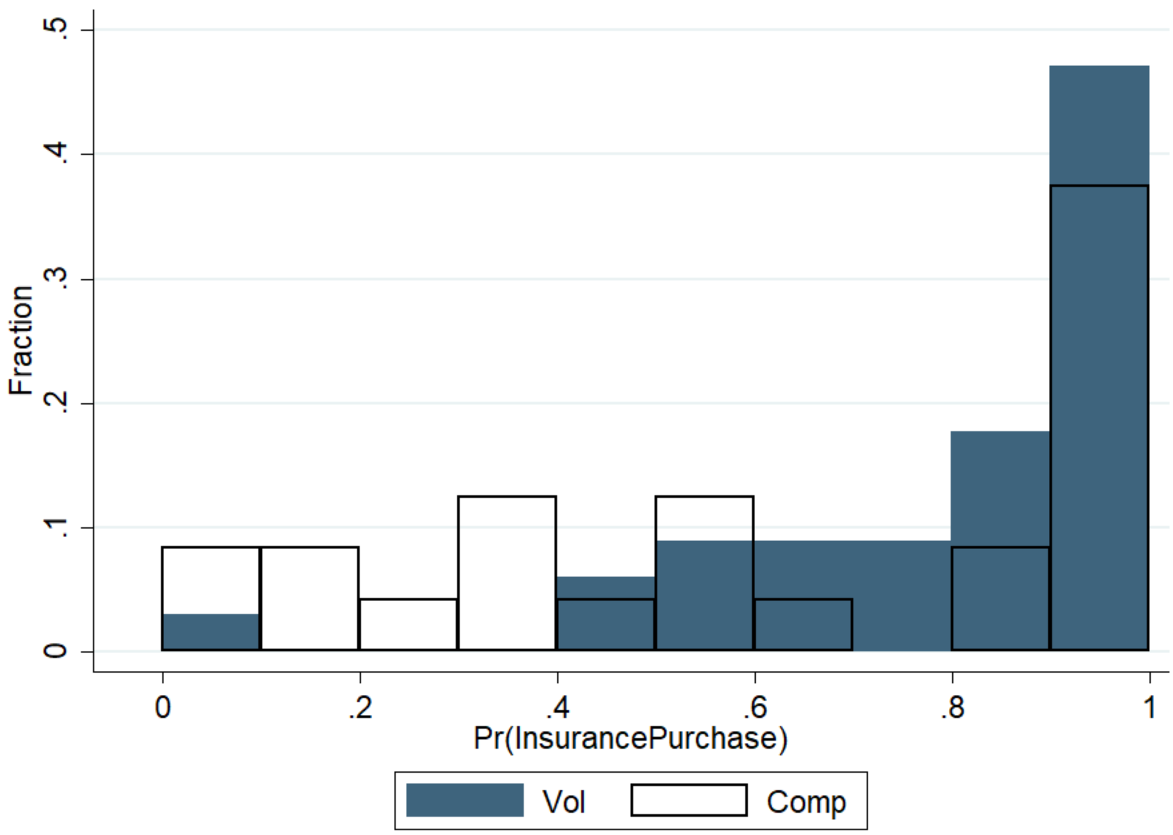


Figure 2.4: Predicted insurance purchase between treatments

The distributions of the predicted insurance probabilities that are depicted in Figure 2.4 differ significantly between the Vol and the Comp treatment (Two-sample Kolmogorov-Smirnov, p = 0.047). While the average predicted insurance probability is 0.81 in Vol it is only 0.61 in Comp with, which suggests that dishonest types self-select into insurance contracts. To control for selection we include the predicted insurance probabilities in our analysis as independent variable and find that the difference in cheating between Vol and Comp turns insignificant (Table 2.2, column 2).[11,12] We conclude that selection effects are the driver of the treatment differences.

[11]This result is confirmed when we perform matching based on the predicted insurance probabilities using a weighted function of the covariates. Results are available from the authors upon request.

[12]It is not feasible to run robustness checks using the observations from the Nudge treatment to derive insurance purchase probabilities, as in that case the selected model has a very low predictive power with a Pseudo R^2 of 0.05 only. Selection effects could thus not sufficiently be controlled for.

VARIABLES	Cheat Standard	Cheat Controlled
Compulsory treatment	-0.23**	-0.10
	(0.11)	(0.11)
Pr(InsurancePurchase)		0.27**
		(0.11)
Loss size	-0.04	-0.04
	(0.03)	(0.03)
Observations	42	42
Pseudo R-squared	0.205	0.349

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2.2: Marginal effects of logit regression on cheating behavior

Result 2.2. *The extent of cheating among individuals that exaggerate their insurance claims is not influenced by the insurance contract formation process. Differences in cheating levels are entirely driven by selection effects.*

2.4.3 Insurer Profitability

Due to the high extent of cheating it is not profitable for the insurer to offer the insurance contract, irrespective of how the insurance contract was concluded. Figure 2.5 shows the average loss per policy holder that an insurer incurs in the different settings. We observe that the insurer's loss is significantly lower when insurance purchase is compulsory as compared to when people are free to choose their coverage (Mann-Whitney tests, $p \leq 0.0158$).

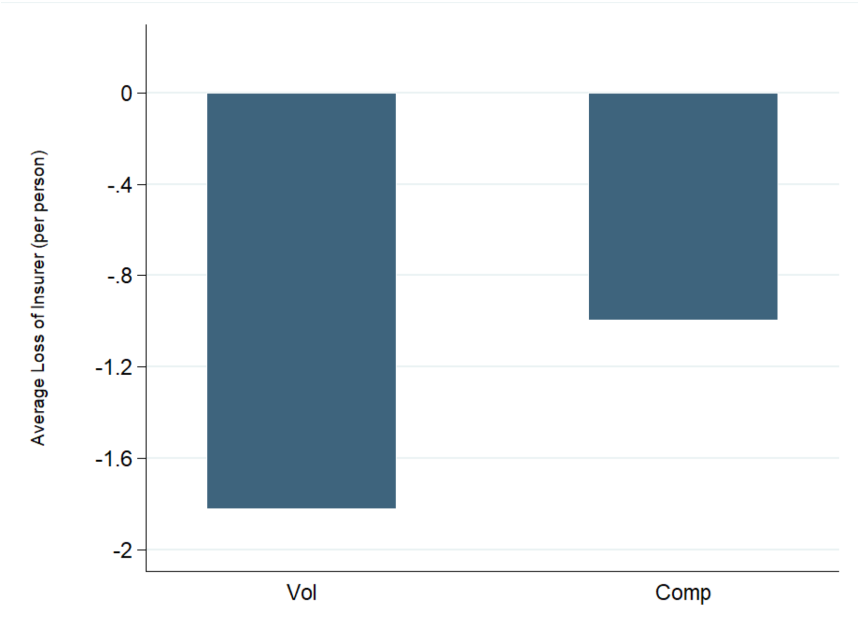


Figure 2.5: Insurer loss by treatment

2.5 Discussion and Conclusion

In this chapter we investigate how allowing people to freely decide whether they want to be covered by an insurance affects insurance fraud. In a laboratory setting we compare loss claiming behavior between settings with voluntary versus mandatory insurance take-up. We find that fraudulent behavior is significantly less pronounced when individuals are required to purchase insurance. The difference can be fully attributed to the self-selection of dishonest subjects into the insurance contract. People who are (not) reluctant to misreport the size of the loss are (more) less likely to join the insurance scheme when given the choice. Under compulsory insurance around 37% of the policy holders behave honestly, while the percentage of non-cheaters is much lower with around 4% among those that self-selected into the insurance contract. Generally, the cheating levels we observe in our insurance setting are rather high, which speaks for a low average level of lying aversion among participants. This may also explain why the risk aversion measure is not a good predictor of the likelihood that participants purchase the insurance.

Our results suggest that incentive schemes and promising contract design may have different or less positive effects respectively, when applied in an environment where adverse selection occurs. For instance, while people who voluntarily select in an insurance with the intention to cheat are may be mainly concerned with the monetary loss they incur when caught, people who select into the insurance because of their risk preferences without an intention to cheat may be relatively more concerned with the psychological costs they expect when considering to cheat. Which measure is most effective in tackling fraud may thus crucially depend on the types of policy holders targeted. To the best of our knowledge, all laboratory economic experiments that investigate solution concepts for ex post moral hazard use a compulsory insurance format (Lammers and Schiller, 2010; von Bieberstein and Schiller, 2017) or offer an actuarially fair premium resulting in (almost) everyone purchasing insurance (Gabaldón et al., 2014). Our findings suggest that it may be useful to test the robustness of their results for voluntary insurances schemes that are priced with a mark-up, as the pool of policy holders and thus the effectiveness of the interventions might differ.

While the results of this chapter speak for the effectiveness of compulsory insurance schemes in tackling ex post moral hazard, there is concern that ex ante moral hazard might be worse under compulsory schemes, especially when the insurance premium is uniform (Chen and Chen, 2013). For example, Cohen and Dehejia (2004) report that the introduction of compulsory automobile insurance comes along with an increase in traffic fatalities. In case of a voluntary insurance setting the fulfillment of protective measures may be introduced as a necessary condition for the conclusion of the insurance contract. The implementation of such requirements is however not feasible in a compulsory setting. While compulsory insurance may thus decrease the precautionary effort in the population, it erases adverse selection in two dimensions. The risk of actual losses is balanced across policy holders, and insurance fraud is mitigated. Both features foster the sustainability of risk sharing arrangements. How the opposing effects of introducing a compulsory insurance scheme eventually balance out in a particular context is an empirical question.

2.A Instructions

General Instructions for Participants

You are about to take part in an economic experiment. The experiment consists of two independent parts. You can earn money in both parts. Your payment depends on your decisions and on chance. At the end of the experiment you will also be asked to fill in a brief questionnaire. In addition, you will receive a flat sum of 4 euro for participating. The money you earn will be paid to you in cash.

Please read the following instructions carefully. You will initially be asked to answer a series of control questions on both parts of the experiment. Only once all participants have correctly answered these questions will we proceed with the experiment.

Communication is prohibited during the experiment. Disobeying this rule will lead to exclusion from the experiment and all payments. If you have any questions, please ask us. Raise your hand and we will come to you.

Information on Part 1 of the Experiment

In this first part of the experiment, we ask you to solve a task. The instructions for this task are on your screen. You are given an income of 3 euro in order to complete the task, in the form of 100 packages, which you will see on your screen. Each of these packages is thus worth 3 eurocent.

The computer will throw away one package per second for you. For each discarded package, 3 eurocent is subtracted from your income. The computer will begin in the top left corner. As soon as a package has been discarded, it will disappear from your screen.

During the experiment, you will always be able to see your current losses in relation to your income. Initially, however, these losses are purely hypothetical, for one of the packages contains a mine that can destroy all other packages. You are not aware which of the packages contains the mine. The mine can be in any of the packages, with the same probability.

It is now your task to stop the computer once you think it has discarded enough packages. Your payment for the first part of the experiment depends on the number of discarded packages and on whether the package containing the mine has been thrown away:

1. If you have thrown away the package containing the mine, you will receive 3 eurocent for each package that has not been discarded.

2. If you have not thrown away the package containing the mine, you will receive 0 euro, because the mine will destroy all packages in your possession.

Please look at the following screenshots. As soon as the 25-second countdown has elapsed, the computer will start to discard the packages one by one. Each corresponding square will turn light gray once the package is gone. The number of discarded and remaining packages will be shown to you in the information field. In addition, you will see the current sum subtracted from your 3-euro income, incurred by the discarding of packages.

If, for example, 2 packages are discarded, the subtracted sum is 2*0.03 = 0.06 euro. If the mine is in one of the 2 discarded packages, your income will be 3 euro 0.06 euro = 2.94 euro. If, say, 98 packages are discarded, the subtracted sum is 98*0.03 = 2.94. If the mine is in one of the 98 discarded packages, your income will be 3 euro 2.94 euro = 0.06 euro. If the mine is in one of the remaining packages, your income will always be 0. To end the discarding of packages, please click the 'STOP' button.

Only at the end of the experiment will you be told in which package the mine was, and informed about the payment resulting from your decision.

Information on Part 2 of the Experiment

In the second part of the experiment, we ask you once again to solve a task. The instructions for this task are on your screen. You are given an income of 11 euro in order to complete the task.

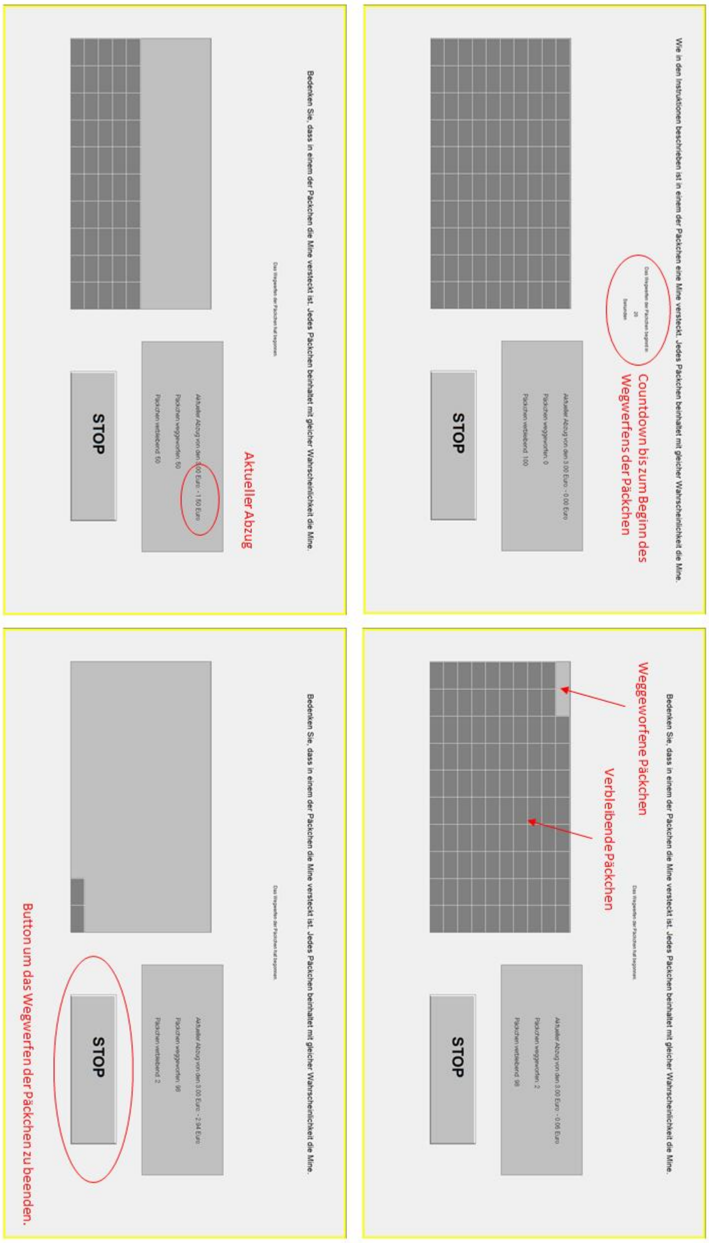
Your income is exposed to risk as the experiment continues. With a probability of 1/3, you will lose 6 euro of your income; with a probability of 1/3, you will lose 3 euro; and with a probability of 1/3, you will lose nothing and keep your entire income of 11 euro.

You are obliged to insure yourself against the risk of loss. This insurance costs 4 euro. It entitles you to reimbursement of the sum you may lose.

At the end of the experiment, you can put in a claim with your insurance. If you claim a loss of 3 euro, 3 euro will be paid to you. If you claim a loss of 6 euro, 6 euro will be paid to you. If you claim no loss, nothing will be paid to you.

Insurance

(Comp) A team member will come to your booth and hand you an envelope marked "Insurance". This envelope contains three boxes. The boxes contain various sums of money corresponding to your income from the task, minus the respective loss and minus the price for the insurance: 11-6-4=1 euro in case of a loss of 6 euro, 11-3-4=4 euro in case of a loss of 3 euro, and 11-4=7 euro in case of no loss.


Wie in den Instruktionen beschrieben ist in einem der Päckchen eine Mine versteckt. Jedes Päckchen beinhaltet mit gleicher Wahrscheinlichkeit die Mine.
Countdown bis zum Beginn des Wegwerfens der Päckchen
STOP
Bedenken Sie, dass in einem der Päckchen die Mine versteckt ist. Jedes Päckchen beinhaltet mit gleicher Wahrscheinlichkeit die Mine.
Aktueller Abzug
STOP
Bedenken Sie, dass in einem der Päckchen die Mine versteckt ist. Jedes Päckchen beinhaltet mit gleicher Wahrscheinlichkeit die Mine.
Weggeworfene Päckchen
Verbleibende Päckchen
STOP
Bedenken Sie, dass in einem der Päckchen die Mine versteckt ist. Jedes Päckchen beinhaltet mit gleicher Wahrscheinlichkeit die Mine.
STOP
Button um das Wegwerfen der Päckchen zu beenden.

(Nudge) There is an envelope marked "Insurance" in your booth. This envelope contains three boxes. The boxes contain various sums of money corresponding to your income from the task, minus the respective loss and minus the price for the insurance: 11-6-4=1 euro in case of a loss of 6 euro, 11-3-4=4 euro in case of a loss of 3 euro, and 11-4=7 euro in case of no loss.

You may choose whether you wish to keep the envelope marked "Insurance", or whether to exchange it for an envelope marked "No Insurance".

The boxes in the envelope marked "No Insurance" contain various sums of money corresponding to your income from the task, minus the respective possible loss: 11-6=5 euro in case of a loss of 6 euro, 11-3=8 euro in case of a loss of 3 Euro, and 11 Euro in case of no loss.

If you wish to buy insurance, please choose the envelope marked "Insurance". If you do not wish to buy insurance, please choose the envelope marked "No Insurance". The insurance decision is initially asked by the computer. Then, depending on your decision, your screen will either show the words "Insurance" or "No Insurance". A team member will come to your booth and exchange the envelope, if appropriate.

(VOL) You may choose between two envelopes. One is marked "Insurance", and the other is marked "No Insurance". Each of these envelopes contains three boxes.

The matchboxes in the envelope marked "No Insurance" contain various sums of money corresponding to your income from the task, minus the respective possible loss: 11-6=5 euro in case of a loss of 6 euro, 11-3=8 euro in case of a loss of 3 Euro, and 11 Euro in case of no loss.

The matchboxes in the envelope marked "Insurance" contain various sums of money corresponding to your income from the task, minus the respective loss and minus the price for the insurance: 11-6-4=1 euro in case of a loss of 6 euro, 11-3-4=4 euro in case of a loss of 3 euro, and 11-4=7 euro in case of no loss.

If you wish to buy insurance, please choose the envelope marked "Insurance". If you do not wish to buy insurance, please choose the envelope marked "No Insurance". The insurance decision is initially asked by the computer. Then, depending on your decision, your screen will either show the words "Insurance" or "No Insurance". A team member will come to your booth and exchange the envelope, if appropriate.

(ALL) Please open the envelope and choose one of the boxes. Please leave the other two boxes unopened in your booth. Open the box you have chosen and take the money contained therein.

Opening more than one box will lead to exclusion from the experiment and all payments.

Claim

Please announce which claim you wish to put in with the insurance. The sum you mention will be covered completely by the insurance and paid to you at the end of the experiment. If you claim 6 euro, you will receive 6 euro at the end of the experiment; if you claim 3 euro, you will receive 3 euro at the end of the experiment; and if you claim nothing, you will receive nothing.

2.B Questionnaire

Demographics

- What is your gender?
- How old are you?
- How tall are you?
- What is your highest degree (Abitur, bachelor, master, doctor, ...)?
- Which subject do you study?
- What is you relationship status (single, in a relationship, engaged, married)?
- What is the highest degree of your mother (Abitur, bachelor, master, doctor, ...)?
- What is the highest degree of your father (Abitur, bachelor, master, doctor, ...)?
- How many siblings do you have?
- What religion do you belong to?
- How often do you consume alcohol?
- How often do you consume cigarettes?

Lying and norm-violating behavior

- Did you ever lie to sell something?
- Did you ever lie in an application for work, a membership, school, university or foundation?
- How often did you go by bike with more than 0.5 per mill?

- How often did you go by car with more than 0.5 per mill?
- Please indicate your estimate how many percent of the other insured participants that are present made a claim that is higher than their actual loss (0-10%, ..., 90-100%)?
- Please indicate to what extent the following statements are true for you (true, rather true, partly, rather not true, not true):
  - I am more likely to lie if there is a lot to win.
  - I am more likely to lie if the chance of being caught is low.
  - Either you lie or you do not lie. There are no further distinctions.

(Soft) paternalism

- Please indicate to what extent the following statements are true for you (true, rather true, partly, rather not true, not true):
  - When I have to make a decision I usually ask for a second opinion.
  - I myself know best what is good for me.
  - I do not always do what is best for me.
- Please indicate to what extent you agree with the following statements (agree, rather agree, neither agree nor disagree, rather disagree, disagree):
  - The government should help smokers who want to quit smoking.
  - Cigarette packages should warn of the detrimental effects of smoking (e.g. via text messages or deterrent pictures).
  - Cigarettes should be taxed.
  - Car drivers should decide for themselves whether to buckle their seat belt.
  - Bicyclists should decide for themselves whether to wear their helmet.

2.C Additional Informations

2.C.1 Insurance Purchase Prediction

	Insurance Purchase
Lied in application	-2.168
	(-1.94)
Male	1.949*
	(2.09)
Age	-0.238*
	(-2.03)
Lied in selling	2.396*
	(2.14)
Belief: other misreport	0.434*
	(2.24)
Constant	2.499
	(0.91)
Observations	51
Pseudo R-Squared	0.4012

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.3: Suggested model for insurance purchase prediction by stepwise command

2.C.2 Loss by Treatment

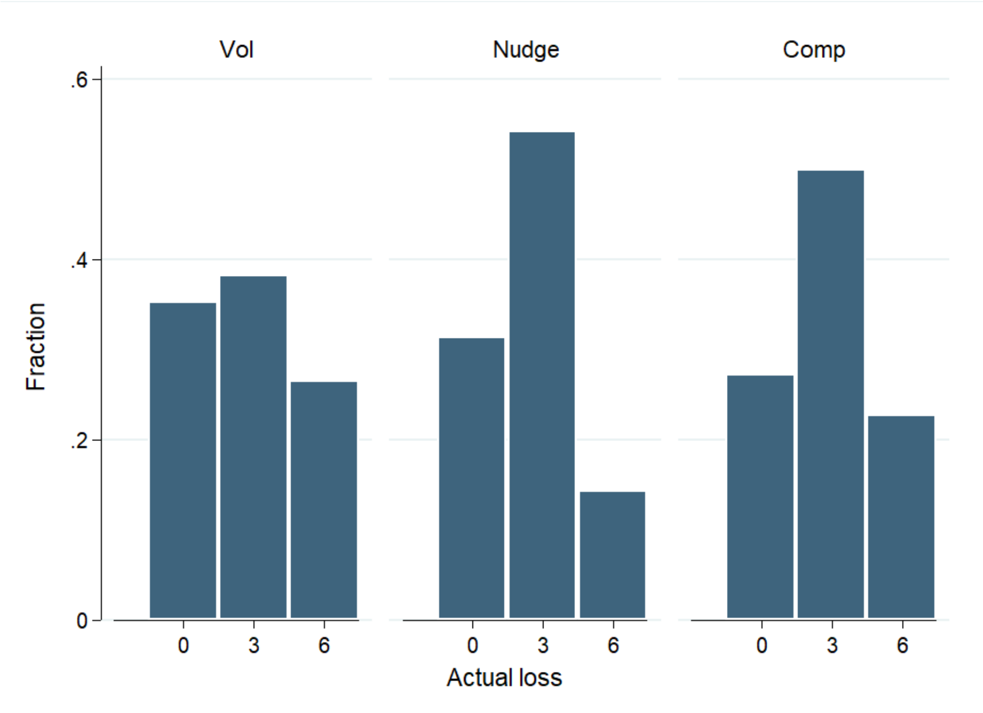


Figure 2.6: The actual losses subjects draw in each treatment

Chapter 3

The Role of Intention in Bilateral Trade Environments: An Experiment

3.1 Introduction

The mechanism design literature provides a powerful set of tools to study the optimal design of institutions. The institutions considered by this literature are able to improve allocations in environments where agents possess private payoff-relevant information. An institution, typically modeled with a social choice function, modifies the framework conditions. A social choice function is then implementable by a mechanism if the desired behavior is in the self-interest of the participating agents.

A useful equivalence result provided by the mechanism design literature is the revelation principle (Myerson, 1981). It states that it is sufficient to restrict the analysis to direct mechanisms that induce market participants to reveal their private information. The advantage of the revelation principle is that it already provides the message set to characterize optimal mechanisms. In this paper we focus on the example of the bilateral trade environment (Myerson and Satterthwaite, 1983). It models a situation where a seller has private information about production costs and is uninformed of the valuation a potential buyer assigns to the produced good. The buyer is uniformed about the seller's production costs.[1]

An insight that the literature provides using the revelation principle in the bilateral trade environment is the Myerson-Satterthwaite Theorem. The theorem states

[1]In general, this environment is a modification of one which illustrates the provision of a public good. But as we conduct a framed experiment, we to generalize our results beyond the bilateral trade environments is difficult to justify.

that it is impossible to implement a social choice function that satisfies incentive compatibility, voluntary participation and budget balance at the same time (Myerson and Satterthwaite, 1983). As an alternative, the authors suggest a minimal subsidy social choice function that violates the budget balance with the smallest possible amount but ensures the other two properties.

In most results provided by the literature agents' preferences are represented by a quasi-linear utility function and thereby create a situation where the Gibbard-Satterthwaite Theorem does not hold (Gibbard, 1973; Satterthwaite, 1975). But a quasi-linear utility function is unable to explain observations made in the behavioral economic literature on an individual level. For instance, Köszegi (2014) provides an overview over the effects of utility functions incorporating behavioral findings in mechanism design and contract theory.

One of the major findings in the behavioral literature is that people care about the intention and payoffs of other individuals (Rabin, 1993; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000). These findings are mainly experimentally investigated in game theoretical environments (Güth et al., 1982) and incorporated in psychological game theory (Geanakoplos et al., 1989). Exactly these strategic considerations between market participants are also applied for the design of mechanisms. In general, findings in behavioral economics contradict the basic idea of mechanism design that the implementation of a social choice function is successful with a mechanism such that the efficient outcome is achieved by the pure self-interests of agents. Hence the question concerning the applicability of mechanisms, that are originally designed under the self-interest assumption, arises in the presence of behavioral effects. For instance, Kucuksenel (2012) shows that with pro-social preferences it is possible to implement more efficient social choice functions. The (positive) incorporation of other individuals' payoffs already aligns partly the objective of the social planner and the agents.

In their paper, Bierbrauer and Netzer (2016) incorporate the idea that agents do not only behave in their self-interest. But in contrast to outcome-based social preferences, the authors combine the idea of intention-based social preference with the mechanism design literature. While in the first part of their paper they explicitly seek for robustness towards all kind of social preferences, the authors study in the second part the influence of intentions on the implementation of social choice functions. Given that reciprocal preference parameters are publicly known, there exists a counterexample in the bilateral trade environment to the revelation principle and the Myerson-Satterthwaite Theorem. This counterexample is a social choice function that shares the trade of gains equally among the two market participants and is only implementable by an indirect mechanism.

In this chapter, we empirically test the provided counterexample and its performance against other mechanisms in the same environment. Our first question concerns whether there exists empirical support for the violation of the revelation principle. We therefore test the direct against the indirect mechanism of the equal share of trades social choice function. In a second step, we implement the minimal subsidy social choice function suggested by Myerson and Satterthwaite (1983) which satisfies robustness towards intention-based social preferences. In the last step, we compare the effect of the outcome of the different mechanisms and the played strategies on the subjective well-being of experimental subjects.

In contrast to the initial hypothesis, we observe the lowest truth-telling rate among all tested mechanisms in the counterexample. Although the possibility to be unkind increases the kindness rating of the truth-telling strategy as predicted by the theory, at the same time it decreases the beliefs that the strategy is played. We also do not find explanatory power of our reciprocal measure on the likelihood that the truth-telling equilibrium strategy is played, but the small number of subjects that play the equilibrium strategy could cause this observation. On the other hand, with respect to the efficient allocations we do not find a difference between the two mechanisms.

In the equilibrium analysis, we conclude that in the direct mechanism of the minimal subsidy social choice function the truth-telling strategy profile forms an equilibrium for any intensity of reciprocal preferences. Empirically, we find that there is indeed an increase in the truth-telling rate in comparison to the indirect mechanism. This provides additional evidence for the importance of robustness concerns in the mechanism design literature. Again, with respect to the efficient allocations we do not find a difference to the counterexample. Note that the indirect mechanism is at a reduced rate for the designer than the minimal subsidy one. With respect to subjective well-being of agents we also do not find any influence of the mechanism on agents that report truthfully. There is also no indication that the procedural fairness of the social choice function affects subjects' well-being.

In the next section we introduce the theoretical framework. We describe the environment that is used in the experimental design and derive our hypotheses for the result section based on the analysis provided by Bierbrauer and Netzer (2016). Then we state in detail how we test the hypotheses in the experiment and the procedure. Afterwards we present the experimental results. Then we discuss potential limitation of the experiment we provide here. In the last section we conclude.

3.2 Theoretical Background

We introduce now the theoretical framework considered for the controlled laboratory experiment. Recalling insights by Bierbrauer and Netzer (2016), we introduce their Bayes Nash Fairness Equilibrium and comment on its relation to the Bayes Nash Equilibrium. Then we use the two equilibrium concepts and the environment to generate hypotheses which we test in the laboratory.

3.2.1 Environment

We focus on an environment which consists of two agents $I = \{b, s\}$ where a seller s produces a single unit good $q \in \{0, 1\}$ for a buyer b. Both, the seller's marginal cost θ_s and the buyer's marginal valuation θ_b are private information and assumed to be either low or high, i.e. $\theta_i \in \Theta_i = \{\underline{\theta}_i, \overline{\theta}_i\}$ for $i \in \{b, s\}$. Both types are equally likely and independently drawn for each agent, i.e. $p(\theta_i) = 0.5 \ \forall \ \theta_i \in \Theta_i, \forall \ i \in I$. We denote the total costs of the seller by $v_s(q, \theta_s) = -\theta_s q$ and by $v_b(q, \theta_b) = \theta_b q$ the total valuation of the buyer. We consider an order of marginal costs and valuations that demands non-trivial solutions, i.e. $0 \leq \underline{\theta}_s < \underline{\theta}_b < \overline{\theta}_s < \overline{\theta}_b$.[2]

A social choice function (SCF) $f : \Theta_b \times \Theta_s \rightarrow A$ specifies for each realization of marginal costs and valuations, $\theta = (\theta_b, \theta_s) \in \Theta_b \times \Theta_s = \Theta$, an allocation. The allocation contains whether the good is produced $q^f : \theta \rightarrow \{0, 1\}$ and the material transfers $t_i^f : \theta \rightarrow \mathbb{R}$ for each agent to the mechanism. We denote also $f = (q^f(\theta_b, \theta_s), t_s^f(\theta_b, \theta_s), t_b^f(\theta_b, \theta_s))$.

In this environment material Pareto efficiency requires that

$$q^f(\theta_b, \theta_s) = \begin{cases} 0, & \text{if } (\theta_b, \theta_s) = (\underline{\theta}_b, \overline{\theta}_s), \text{ and} \\ 1, & \text{otherwise.} \end{cases}$$

A mechanism Γ contains of a message set M_i for each agent and an outcome function $g : M_b \times M_s \rightarrow A$, i.e. $\Gamma = [M_b, M_s, g]$. We denote similar to social choice functions f an outcome function as $g = (q^g(\theta), t_b^g(\theta), t_s^g(\theta))$.

A pure strategy for agent i in mechanism Γ is a function $s_i : \Theta_i \rightarrow M_i$. The set of strategies for both agents is denoted $S = S_b \times S_s$. In addition, we introduce first and second order beliefs. As the analysis of Bierbrauer and Netzer (2016) focuses on

[2]Note that this environment is equivalent to one of the efficient public good provision where one agent's gains from the provision while another looses. When the magnitude of gains and losses are private information an efficient social choice functions provides the good only if gains outweigh losses.

pure strategy equilibrium we consider only beliefs with a unit mass[3]. With $s_i^b(\theta_j)$ we denote the (first order) belief of agent i about $j's$ pure strategy for private type θ_j. Similarly, $s_i^{bb}(\theta_i)$ denotes the (second order) belief of agent i about $j's$ belief about agent $i's$ pure strategy given realization θ_i.

Assume mechanism Γ, then agent i's strategy s_i given the belief s_i^b yields an ex ante expected payoff of

$$\pi_i(s_i, s_i^b) = \mathbb{E}_\theta\big[v_i(q_i^g(s_i(\theta_i), s_i^b(\theta_j)), \theta_i) + t_i^g(s_i(\theta_i), s_i^b(\theta_j))\big] \tag{3.1}$$

where the expectation is taken over all possible combinations of private informations. A strategy profile that yields the highest ex ante expected payoff given correct first order beliefs for both agents forms a Bayes Nash equilibrium. Formally, we define that

Definition 3.1. *A Bayes Nash Equilibrium is a strategy profile $s^* = (s_b^*, s_s^*)$ such that, for both agents $i \in \{b, s\}$,*

(a) $s_i^* \in argmax_{s_i \in S_i}\ \pi_i(s_i, s_i^b)$,

(b) $s_i^b = s_j^*$.

In the next step the payoff is expanded by reciprocal concerns and we introduce the new solution concept: a Bayesian form of the model by Rabin (1993). In detail, given a strategy plan, first and second order beliefs, the ex ante expected payoff is compared to an equitable reference payoff which we denote by $\pi_j^e(s_i^b)$. In Bierbrauer and Netzer (2016) Rabin's original composition of the equitable reference payoff is adapted to Bayesian' environments such that we set

$$\pi_j^e(s_i^b) = \frac{1}{2}\big[\max_{s_i \in E_i(s_j^b)} \pi_j(s_i^b, s_i) + \min_{s_i \in E_{ij}(s_j^b)} \pi_j(s_i^b, s_i)\big]$$

where $E_i(s_j^b)$ denotes the set of all strategies that yield Pareto efficient outcomes. We use $\pi_i^e(s_i^{bb})$ to define kindness and to derive our hypothesis. Whether two subjects homogeneously define an action either kind or unkind remains an open empirical questions and other formations have been proposed in the literature (see for instance, Dufwenberg and Kirchsteiger, 2004; Çelen et al., 2017).

Given these equitable reference payoffs, we next define two kindness terms. The

[3]In the experiment we restrict the choice set of beliefs to a unit mass. But note that it is impossible to restrict the true underlying beliefs of experimental subjects to the unit mass and it is therefore only an estimate for the true belief distribution.

first one is how kind i expects to be towards j, i.e.

$$\kappa_i(s_i, s_i^b) = \pi_j(s_i^b, s_i) - \pi_j^e(s_i^b)$$

which compares the ex ante expected payoff to agent j given strategies s_i and s_i^b in comparison to the equitable reference payoff given s_i^b.

The second one is the expectation of i about how kind j expects to be to i which we denote by

$$\kappa_i(s_i^b, s_i^{bb}) = \pi_i(s_i^{bb}, s_i^b) - \pi_i^e(s_i^{bb}).$$

Thereby the ex ante expected payoff that agent i would receive given s_i^b and s_i^{bb} is compared to the equitable reference payoff i beliefs j beliefs to be fair.

In total, the ex ante expected utility is given by the sum of the ex ante expected payoff and the expected bilateral kindness whereby the later is weighted by the reciprocity parameter γ, such that

$$U_i(s_i, s_i^b, s_i^{bb}) = \pi_i(s_i, s_i^b) + \ \gamma\ \kappa_i(s_i, s_i^b)\ \kappa_i(s_i^b, s_i^{bb}). \tag{3.2}$$

Given the ex ante expected utility function, we modify the BNE to the BNFE. A strategy profile that yields the highest ex ante expected utility given correct first and second order beliefs for both agents forms a Bayes Nash Fairness Equilibrium. Again, formally, we define that

Definition 3.2. *A Bayes Nash Fairness Equilibrium is a strategy profile $s^* = (s_b^*, s_s^*)$ such that, for both agents $i \in \{b, s\}$,*

(a) $s_i^* \in argmax_{s_i \in S_i}\ U_i(s_i, s_i^b, s_i^{bb})$,

(b) $s_i^b = s_j^*$,

(c) $s_i^{bb} = s_i^*$.

3.2.2 Analysis

We continue with the theoretical insights which we investigate with our experimental design and derive testable hypotheses for the laboratory. The valuation parameterizations $\underline{\theta}_s = 0$, $\underline{\theta}_b = 20$, $\overline{\theta}_s = 80$, $\overline{\theta}_b = 100$ implements a discrete version of the environment where the Myerson-Satterthwaite Impossibility Theorem (Myerson and Satterthwaite, 1983) applies. The theorem states that it is impossible to implement a social choice function in a Bayes Nash equilibrium and satisfy at the same time participation constraints as well as (ex post) budget balance.

Table 1 shows the social choice function f^e where the gains from trade are shared equally between the seller and the buyer.

	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(1, 10, -10)$	$(0, 0, 0)$
$\overline{\theta}_b$	$(1, 50, -50)$	$(1, 90, -90)$

Table 3.1: Equal share direct SCF $f^e(\theta_1, \theta_2)$

As the equilibrium analysis in Appendix 3.A shows, the direct mechanism neither implements f^e in a BNE nor in a BNFE for any reciprocal weighting $\gamma > 0$.

In contrast, Bierbrauer and Netzer (2016) show that for any strictly positive reciprocal weight there exists an indirect mechanism (see Table 3.2) that implements f^e in a BNFE. The indirect mechanism includes an additional unkind message $\tilde{\theta}_i$ for each role $i \in \{b, s\}$. The outcome of the message is equal to the weak type ($\underline{\theta}_b$ or $\overline{\theta}_s$) but increases the own share of trade gain by δ_i. For the parameterization $\delta_b = \delta_s = 5$, the truth-telling strategies form a BNFE for a modest range of reciprocal weights, i.e. $\gamma \in [0.2, 0.45]$. The existence of such an equilibrium is a counterexample to the revelation principle.

	$\underline{\theta}_s$	$\overline{\theta}_s$	$\tilde{\theta}_s$
$\tilde{\theta}_b$	$(1, 5, -5)$	$(0, 0, 0)$	$(0, 0, 0)$
$\underline{\theta}_b$	$(1, 10, -10)$	$(0, 0, 0)$	$(0, 0, 0)$
$\overline{\theta}_b$	$(1, 50, -50)$	$(1, 90, -90)$	$(1, 95, -95)$

Table 3.2: Equal share indirect mechanism Γ^e_{ind}

The equilibrium analysis shows that there exist a set of non-truth-telling BNE and BNFE in both induced games. This multiplicity of equilibria raises the empirical question whether the indirect mechanism indeed implements f^e. Our first hypothesis therefore is:

Hypothesis 1a: *The truth-telling strategy is more likely to be played in the game induced by the indirect mechanism than by the one induced by the direct mechanism.*

Investigating the difference between implementation in the direct and indirect mechanism is of interest for two reasons. First, the majority of projects in the mechanism design literature applies the revelation principle and concentrates on implementation in the direct mechanism. The verification that indirect mechanisms lead to a more efficient allocation in environments where social motives are likely to occur, is an empirical indication that additional implementation in indirect mechanism should be considered.[4]

The same argument applies to the empirical verification of results in controlled laboratory experiments. To our knowledge, most experimental projects testing mechanism design theories solely focus on designs where subjects play the game induced by the direct mechanism.

Second, *Hypothesis 1a* implies that the resulting allocation of the indirect mechanism is more efficient. In the truth-telling strategy profile that forms an BNFE trade takes only place when the valuation is larger than the cost. The equilibrium analysis shows that there exists additional equilibria besides the truth-telling one that implement the efficient allocation. But there also exists a set of (unkind) BNFE where no production takes place in the indirect mechanism. On the other hand, we find that with specific inconsistent beliefs the truth-telling strategy is still a best response in the direct mechanism.[5] Therefore also the efficiency of the two mechanism remains an empirical questions and constructs our next hypothesis.

Hypothesis 1b: *The indirect mechanism provides a more efficient allocation than the direct mechanism.*

According to the theory, the strategy chosen by an agent depends crucially on the beliefs about behavior and first order beliefs of their counterparty. This belief structure directly determines which equilibrium is played and whether the allocation is indeed more efficient provided by the rules of the indirect mechanism. As first and second order beliefs are elicited in the experiment, we can also test whether

Hypothesis 1c: *The truth-telling strategy in the game induced by the indirect mechanism is more likely to be played if the reciprocity is modest and the beliefs are such that the other agent's strategy and first order beliefs are given by the truth-telling strategy.*

[4] See for this argument also the work by Kucuksenel (2012).

[5] In detail, there exist two equilibria $(\theta_b^L, \theta_b^H, \theta_s^H, \theta_s^H)$ and $(\theta_b^L, \theta_b^L, \theta_s^L, \theta_s^H)$ where truthful revelation is optimal. If both agent have inconsistent beliefs about the other agent's strategy, the allocation would be efficient although the social function is not implemented.

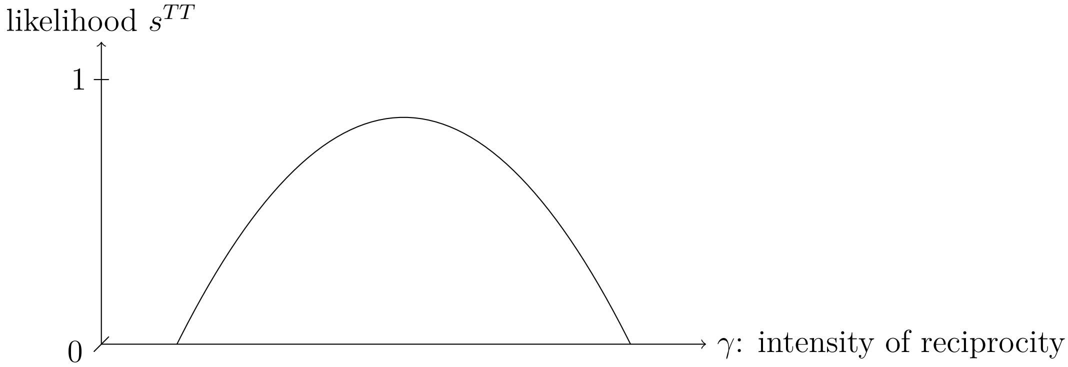


Figure 3.1: Predicted effect of reciprocity on the likelihood of truth-telling

In the case the relative weighting of reciprocal motives to material gains is too large, the truth-telling strategy profile is no longer a BNFE. Instead there exists the inefficient (kind) BNFE $(\theta_s^L, \theta_s^L, \theta_b^H, \theta_b^H)$ where the good is always produced. This includes the inefficient case if the producer's costs exceed the buyer's valuation, i.e. $\theta_s > \theta_b$.

One advantage of a framed experiment is that different roles are randomly allocated to experimental subjects. This random assignment allows for causal inference. The question whether the role in an (artificial) society affects behavior and beliefs are expected can additionally be tested. As the theory does not predict different behavior, we state the two-sided hypothesis that

Hypothesis 2: *(a) First and (b) second order beliefs about the counterparty differ between subjects in the role as seller and as buyer.*

Myerson and Satterthwaite (1983) themselves also provide a social choice function for the bilateral trade environment. The minimal subsidy social choice function f^* (see Table 3.3) violates the ex post budget balance by the minimal possible amount, but participation constraint and Bayesian incentive compatibility are satisfied. In the experiment we double the minimal subsidy for each trade from five to ten in order to simplify the experimental parameterization for subjects. In the equilibrium analysis of the game induced by this social choice function f^*, we conclude that the truth-telling strategy profile forms a BNE and for any weighting $\gamma > 0$ also a BNFE.

	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(1, 30, -25)$	$(0, 5, 5)$
$\overline{\theta}_b$	$(1, 55, -45)$	$(1, 80, -70)$

Table 3.3: Minimal subsidy SCF $f^*(\theta)$

The reason that the BNFE exists for any γ is that the truth-telling strategy of both agents are perceived as neither kind nor unkind. Hence a rewarding deviation needs to increase the material payoff which is excluded because the truth-telling strategy profile forms a BNE.

There are two questions arising when comparing the BNFE in the minimal subsidy direct mechanism to the BNFE formed in the indirect mechanism for the equal share social choice function: The robustness of the mechanism as well as the happiness of the subjects.

First, the subsidy mechanism is robust against arbitrary reciprocal weighting parameters. Indeed, the truth-telling strategy is the unique equilibrium that exists for all $\gamma \geq 0$. In contrast, the BNFE that implements f^e in the indirect mechanism only exists for a specific range of parameters. Hence, we predict that the likelihood that the strategy is indeed played is larger in the game induced by f^e than in the indirect mechanism. Consequently, our hypothesis is that

Hypothesis 3: *The direct mechanism of the social choice function f^* provides a more efficient allocation than the indirect mechanism of the social choice function f^e.*

Second, if for the implementation f^* no (believed) bilateral kindness is involved, then agents receive utilities exclusively from their material gains. This is different for the implementation of f^e in the indirect mechanism. Next to their material gains agents are assumed to receive in this case additionally (psychological) utility from the (believed) bilateral kindness. The (believed) positive intention could generate additional happiness for the market subjects. From a behavioral welfare economics perspective the question is whether this additional happiness outweighs the robustness tested in the former hypothesis. But for a potential comparison, we first need to test whether

Hypothesis 4a: *Subjects report a higher happiness after using the truth-telling strategy in the indirect mechanism of f^e than in the direct mechanism of f^* and whether this is correlated with reciprocal motives.*

Concerning the welfare of subjects in bilateral trade environments another question rises up: Whether in general the price construction directly influences the happiness of subjects. The two direct mechanisms of the different social choice functions propose different methods to calculate the trading price.

While the first one shares the gains between the two parties, the later lowers the payoff for weak types (low valuation or high costs) in order to guarantee incentive compatibility for the strong types. Hence, weak types are disadvantaged payoff wise, which might be generally considered as an unfair trading rule. In contrast, the first rule might to be perceived as fairer with respect to equality and directly raises happiness. In our last hypothesis we test therefore whether

Hypothesis 4b: *Subjects report a higher happiness in general after they play the game induced by the direct mechanism of* f^e *than by the one of* f^**.*

3.3 Experimental Design

The experimental design can be distinguished chronologically into two parts. The first part takes place online via Qualtrics and allows us to gather individual data points for each subject. We elicit intention and outcome-based social preferences for each subject and use a survey to obtain socio demographic data.

The second part is then executed in the laboratory via zTree (Fischbacher, 2007) where subjects are randomly allocated into one of three treatments. Between subjects they play either the game induced by the direct or the indirect mechanism of the equal share social choice function in the bilateral trade environment. The third treatment is the game induced by the direct mechanism of the minimal subsidy social choice function.

Next to a series of survey questions, we elicit their beliefs about behavior and first order belief of the subject they are assigned to. To avoid hedging strategies, either the outcome of the induced game or the elicitation of the beliefs is paid out in addition to the online part.

3.3.1 Online Experiment

As the main focus of the study is on the effect of reciprocal motives, we gather in a first step individual data on reciprocal behavior in a mini ultimatum game (UG).

We adapt the version by Falk et al. (2003) where subjects play four different versions of the UG (see Figure 3.2).

Subjects are assigned to a partner for this task where one takes the role as the sender and the other as receiver. We apply the strategy method to elicit reciprocal parameters for each subject. Hence, the procedure is the same for all subjects. First, each subject plays as a receiver all versions in a random order. The task as receiver is to accept or reject an offer of the sender about how to split up 100 (online) experimental points by the sender. The sender can send one of two offers. One is always a share of $(20, 80)$. The alternative offer is manipulated between the versions and is either $(20, 80)$, $(50, 50)$, $(80, 20)$ or $(100, 0)$.

Again, we use the strategy method within each version such that the receivers needs to state whether she would accept for each of the two possible offers. Afterwards, each subject also plays each version in the role of a sender.

Reciprocity assumes that the perceived kindness to pick option $(20, 80)$ depends on the alternative offer. Therefore, the acceptance rate of option $(20, 80)$ should change between the versions, if the subject indeed has reciprocal preferences. The fair reference payoff for the receiver increases with the alternatives. Hence, the later a switch from accept to reject in the $(20, 80)$ option, the less the subject is assumed to have reciprocal preferences.[6]

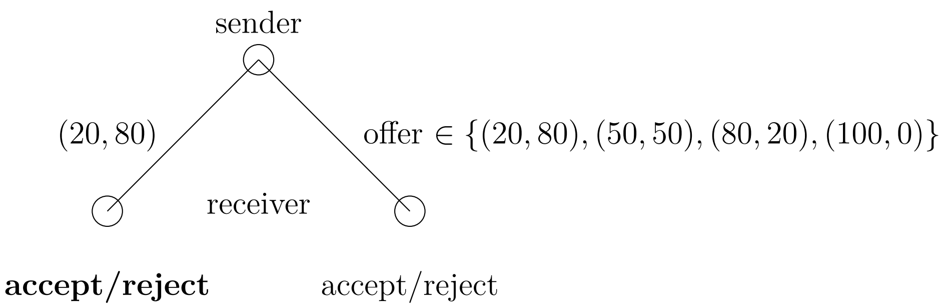


Figure 3.2: Mini ultimatum game

In contrast, under pure outcome-based preferences this acceptance rate should be the same, independent on the alternative offer. To receive more insight into the subjects' outcome-based preference we elicit the SVO angle for each subject using the SVO Slider Task by Murphy et al. (2011).

[6]A disadvantage of this elicitation method is that we measure recirpocity with respect to certain outcomes. In the laboratory experiment later, we assume reciprocity with respect to lotteries, i.e. risky outcomes. The interaction between risk and social preferences is indeed non-trivial (Bolton et al., 2005). The theoretical model we use to derive our hypotheses assumes risk neutrality in material payoffs. Experimental evidence shows though that the majority of experimental subjects behave risk averse even in small stake games.

Further, we ask for their social demographics and additional questions concerning positive and negative reciprocity from the global preference survey (Falk et al., 2016). In Appendix 3.F we provide the detailed formulation of these questions.

3.3.2 Laboratory Experiment

Treatments

In the laboratory part of the experiment each subject is allocated randomly into one of three treatments and is unaware of the existence of the other two. In one session only one treatment is played by all subjects to ensure that they do not receive information about the other treatment conditions. Each treatment represents one of the games induced either by the direct (**BASE**) or indirect mechanism (**MSE**) to implement the equal share social choice function or by the direct (**PM**) to implement the minimal subsidy social choice function.

Before the start of the main experiment, subjects receive instructions for the whole experiment (see Appendix 3.G for details). In order to ensure understanding, they can simulate results by entering different strategies for both roles before the actual experiment starts. Details of the simulation can be found in Appendix 3.E. After the simulation subjects need to correctly answer control questions before the start of the actual main part of the experiment.

Baseline (BASE)

Each subject plays the game once either in the role as seller or buyer and is assigned to another subject with a different role. The game is framed as a trade environment where the seller can produce a good which cost her in points either 0 or 80. Both costs are equally likely. These points will be subtracted from her payoff if there is trade. The buyer values the good in points at either 20 or 100 which are also equally likely. The valuation is added to the payoff in the case that the good is produced. To avoid a loss frame each subject is initially endowed with 100 points and every subject is asked to enter a strategy.

In a strategy a message for each of the two valuation/costs (low and high) needs to be submitted to the mechanism. Hence, subjects play the game in the ex ante stage where they are not aware of neither their own nor the other subject's valuation/costs. The messages for the buyer are labeled as maximal asking price and for the seller as minimal asking price.

The allocation function states that the good is produced if the maximal asking price is larger than the minimal asking price. The price is given by

$$price_{base} = (maximal\ asking\ price + minimal\ asking\ price) : 2.$$

and only needs to be paid from the buyer to the seller if the good is traded.

Message Set Expansion (MSE)

This treatment differs from the baseline by an expansion of the message sets of both roles. Instead of only low and high reports, each role is able to report an additional 'unkind option'. The seller can report a 'very high minimal asking price' of 90 points while the buyer can report 'very low maximal asking price' of 10 points for the good. In technical terms, we set $M_b = \{10, 20, 100\}$ and $M_s = \{0, 80, 90\}$.

Price Modification (PM)

In the minimal subsidy every subject receives 5 points independent whether good is produced or not. The price paid from the buyer to the seller in the case of production is the same as in the first two treatments added up with a *Bonus*:

$$Bonus = \begin{cases} -15 & \text{if minimal asking price} = \text{maximal asking price} = \text{low}, \\ +15 & \text{if minimal asking price} = \text{maximal asking price} = \text{high}, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, in this treatment the price calculation is given by

$$price_{PM} = \pm\ 5 + price_{base} + Bonus.$$

Questionnaire

After subjects report their strategy in the game induced by the respective mechanism, we elicit in each treatment first and second order beliefs. In a first step each of them reports beliefs about the other subject's strategy. One message for the low type, and one for the high one. We use point estimation as the theory is based on unit mass probabilities and it simplifies the questions. With each correct answer they earn 30 points. Each subject is then asked to report their beliefs about the other subject's answer on the first question. These are the second order beliefs. Again, each correct answer is worth 30 points.

In addition, subjects are asked to answer additional questions about the kindness of all possible strategies the assigned subjects have. Further questions include whether they would have joined if entry would have been voluntarily and state what their current level of subjective well-being[7] is. For the reference point by Çelen et al. (2017) subjects report non-incentivized the strategy, they would have played in the not assigned role. In Appendix 3.F we provide the list questions. When all subjects finished the questionnaire, feedback is provided about the outcomes of the online part, the two parts in the laboratory, and which of them is payoff relevant.

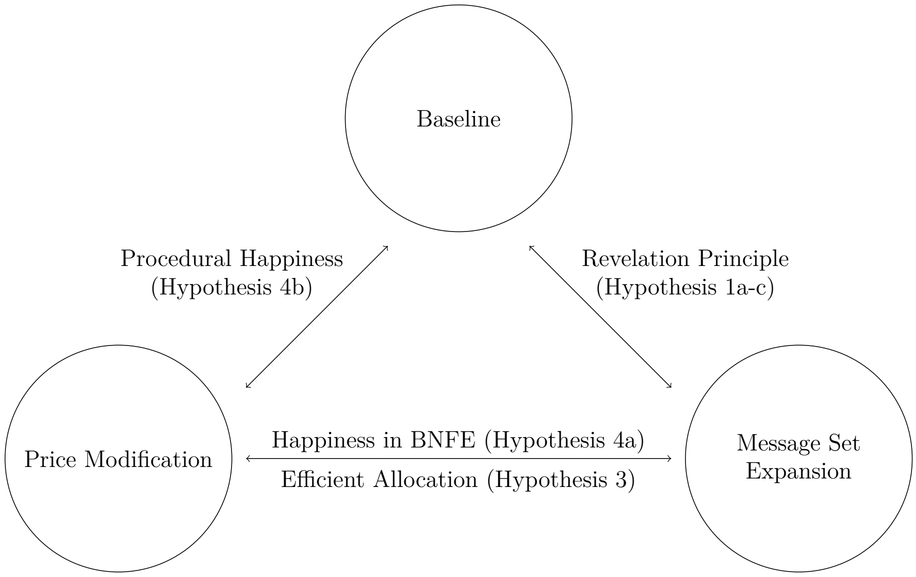


Figure 3.3: Treatment and hypotheses overview

3.3.3 Experimental Procedure

We conducted two sessions for each of the three treatments, with 128 subjects in total. In each of the BASE and the MSE treatment 42 subjects participated and in the PM treatment 44, which varies due to no-shows. Each subject participated only in one session which lasted on average about 55 minutes. Subjects are invited using ORSEE (Greiner, 2015). The online experiment was computerized with the software Qualtrics and the laboratory experiment with z-Tree (Fischbacher, 2007). The experiment took place in the DecisionLab at the Max Planck Institute for Research on Collective Goods, Bonn in May and June 2018. Subjects were mainly students from the University of Bonn.

Participating subjects received an email with the invitation to the online experiments one week before the laboratory experiment. Subjects were informed that they had time to finish the online experiment up to 24 hours before the beginning of the laboratory experiment and that the completion was necessary to take part in the

[7]In the questionnaire we state three questions to measure subjective well-being. We ask for their current mood, general life satisfaction and whether they consider themselves as 'good humans'. For details, see Appendix 3.F.

laboratory experiment. When subjects enter the room, the instruction were already at their cube. The translated instructions are provided in Appendix 3.G. After subjects finished reading them, they were asked to enter their personalized code[8] and could use the simulation to become acquainted with the experiment. When every subject finished the control questions, the actual experiment started.

Subject can earn experimental points which will be exchanged to Euro after they complete the whole experiment. In the online part 50 points were converted into 1 euro while in the laboratory experiment 20 points were worth 1 euro. On average subjects earned 13.50 euro, including a show-up fee of 5 euro.

3.4 Experimental Results

In the first part of this section we state the empirical findings in the experiment with respect to hypotheses derived in the theoretical background in Section 3.2.

Direct vs. Indirect Mechanisms

Basically, the revelation principle states that every social choice function that is implementable in an indirect mechanism needs to be also implementable in the corresponding direct mechanism. This result suffices for the focus on direct mechanisms in the literature.

Bierbrauer and Netzer (2016) conclude that there exists the possibility to implement a social choice function in an indirect mechanism while it remains impossible in a direct one. The theoretical finding requires that the agents reciprocate kind and unkind behavior. An example the authors present is a social choice function which equally shares gains of trade between agents in the bilateral trade environment.

We introduce this counterexample to the revelation principle in detail in Section 3.2.2. In our first hypothesis, we empirically test whether the likelihood of the truth-telling strategy indeed increases, if the implementation is in a indirect mechanism instead of a direct one.

Contrary to the theoretical prediction, we find that

Result 3.1. *The implementation of the equal share of trade gains social choice function is less, not more, likely in the indirect than in the direct mechanism.*

Figure 3.4 shows the frequency that subjects in the experimental treatments play the truth-telling strategy. Remember that theoretically the truth-telling strategies

[8]The personalized code allows us to match the data generated in the online experiment with the one in the laboratory anonymously.

form a BNFE, only if first and second order beliefs are also truthful. On the left side of Figure 3.4 we display only whether the strategy is indeed to reveal the own information while on the right side we condition on truthful first and second order beliefs. For both cases, we conclude that the frequency in the Message Set Expansion treatment (**MSE**) not increases in comparison to the Baseline treatment (**BASE**) ($p = 0.99$, Mann Whitney U Test [one-sided][9]).

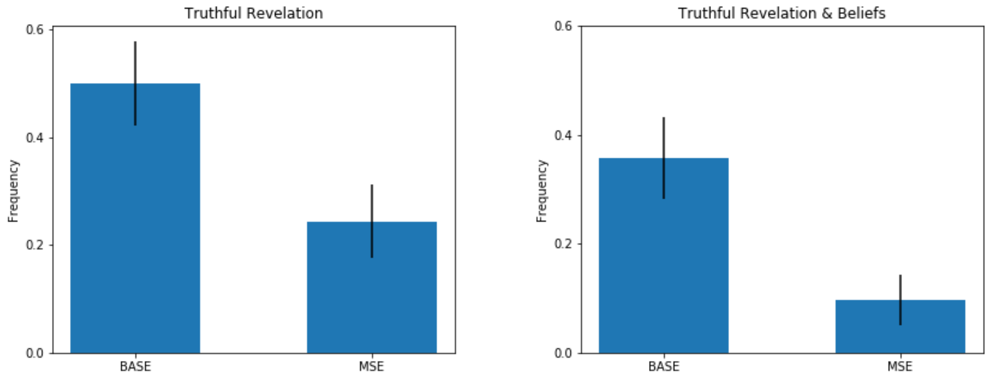


Figure 3.4: Frequency of the truth-telling strategy for the direct and indirect mechanism

In most intention-based social preference models all strategies are evaluated with respect to a social reference point. In the model by Bierbrauer and Netzer (2016) this social reference point depends on the Pareto efficient message set. Hence, whether a certain strategy is perceived as either kind or unkind also depends on the message set provided in the mechanism.

In the environment used in the experimental design, the inclusion of unkind messages increases theoretically the kindness perception of the truth-telling strategy. The expansion of the message set implies that more messages exist than private information types, i.e. it is no longer a direct mechanism. In Figure 3.5 we show empirical support for such a relationship. Indeed, we conclude that the kindness rating of the truth-telling strategy is significantly larger in **MSE** than in **BASE** ($p = 0.036$ Mann Whitney U Test; $p = 0.012$ Kolmogorow-Smirnow-Test).

[9] If we do not state explicit otherwise, we always report the two-sided test statistic or p-value.

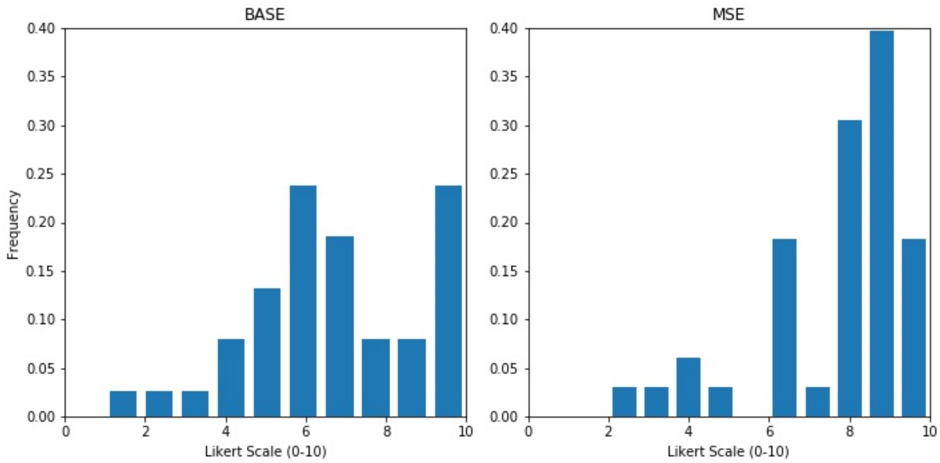


Figure 3.5: Kindness rating of the truth-telling strategy by the other subject

As the kindness perception increases as predicted by the theory, we now test whether the beliefs can account at least partially for our empirical results. Indeed, we find that in **MSE** beliefs decrease significantly that the other subject reports her private information types truthfully. Figure 3.6 illustrates this significant decrease in first order beliefs between the treatments ($p = 0.019$, Mann Whitney U Test). Although the overall levels are large, there is a marginal significant decrease in the second order beliefs between the two treatments ($p = 0.064$, Mann Whitney U Test). Hence, they also anticipate that the other subject is less likely to belief that they report truthful. We return to this finding in detail for the test of *Hypothesis 2*. There we report that the effect is mainly driven by buyers' first order beliefs about sellers' behavior. In **MSE** there are significant differences in first order beliefs, which is not the case in **BASE**. Also, the decrease of buyers' belief between treatments is significant different.

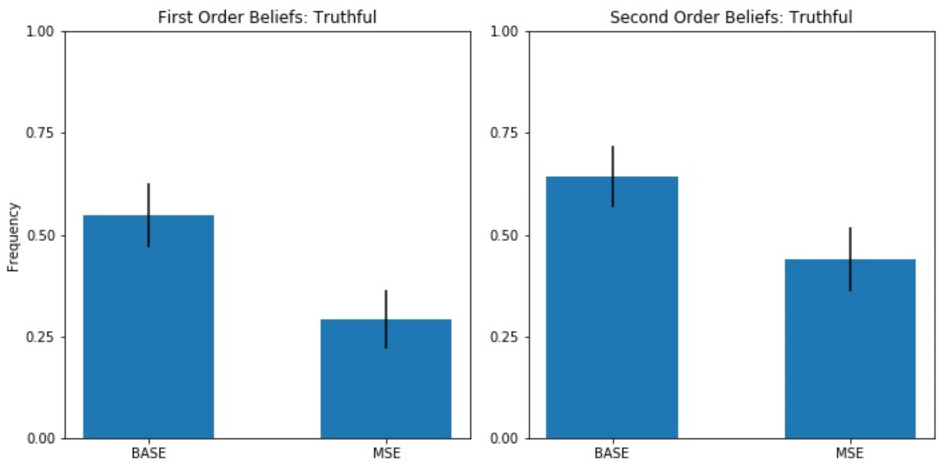


Figure 3.6: First and second order beliefs that the other subject reports truthful

As we already discussed in the theoretical background section, there exist a set of equilibria beside the truth-telling strategy profile in **MSE** treatment that also lead to an efficient allocation. The same is not true for the game in **BASE** where in theory no equilibrium leads to an efficient allocation. There exist a set of equilibria where beliefs are such that the other subject does not truthfully reveal for which the truth-telling strategy is a best response.

As inconsistent beliefs are empirically possible we might observe that the truth-telling strategy is played by both subjects. As the realization of efficient allocations is the main focus in classical welfare economics, we test with our *Hypothesis 1b* whether the efficient trade increases in **MSE** in comparison to **BASE**. As illustrated by Figure 3.7, we find that

Result 3.2. *There is no difference in the allocation efficiency between direct and indirect implementation of the equal of trade gains share social choice functions (p=0.11, Mann Whitney U Test).*

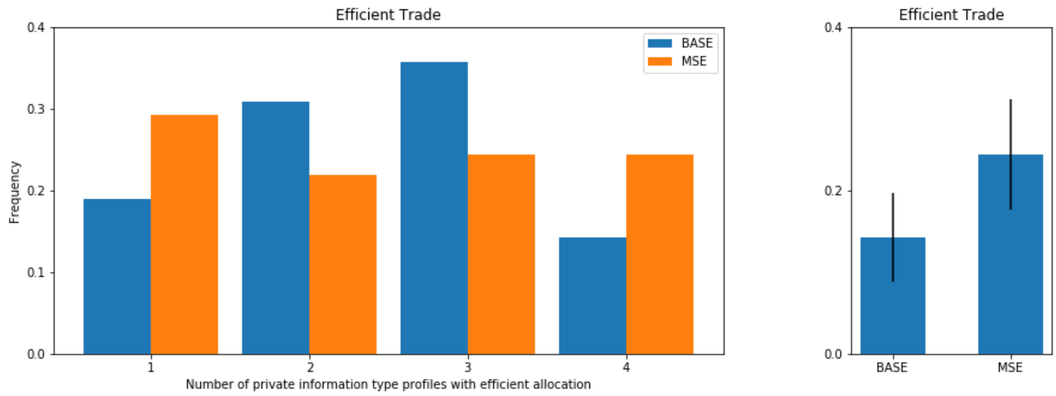


Figure 3.7: Comparison of efficient trade in BASE and MSE: complete distribution & full efficiency

Although there are significant fewer subjects that truthfully reveal in **MSE** than in **BASE**, the multiplicity of equilibria implies no differences in efficient trade. Actually, Figure 3.7 slightly indicates that the indirect mechanism seems to be more efficient in terms of likelihood of full efficiency.

In the test of *Hypothesis 1c* we use the individual data collected in the online experiment. For each subject we collect several measures of reciprocal behavior and the SVO angle. We use these data points in a Probit model to check whether they significantly affect the likelihood that subjects play the truth-telling equilibrium strategy. Table 3.4 reports different model specification.

We focus on the unincentivized measurements for the intensity of individual reciprocity because we only observe small variation in the answers to the incentivized one. There exist two questions that measure positive direct reciprocity which are a self assessment and a scenario where subjects can hypothetically reward kind behavior towards them. The questions are provided in detail in Appendix 3.F.

Table 3.4: The effect of reciprocity on the likelihood of truth-telling strategy in MSE

	measurement: present scenario			measurement: self-assessment		
	(1)	(2)	(3)	(4)	(5)	(6)
	model 1	model 2	model 3	model 1	model 2	model 3
reciprocity	0.1186	0.1011	0.0529	-0.1596	-0.1533	-0.0856
	(0.577)	(0.656)	(0.867)	(0.412)	(0.455)	(0.739)
reciprocity^2	0.0795	0.0920	0.1525	0.017	0.0575	0.1107
	(0.620)	(0.590)	(0.477)	(0.993)	(0.778)	(0.705)
kindness truth-telling		0.2232	0.3359		0.2399	0.2922
		(0.309)	(0.265)		(0.304)	(0.283)
svo angle		-0.0035	-0.0032		0.0004	-0.0022
		(0.879)	(0.909)			(0.943)
sociodemographic			Yes			Yes
constant	-2.2756*	-3.9943*	-8.8663	0.2137	-2.3637	-11.0137
	(0.078)	(0.077)	(0.99)	(0.933)	(0.497)	(0.99)
N	42	42	42	42	42	42

p-value in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table 3.4 we present several Probit model specifications that explain the likelihood that a subject played the truth-telling equilibrium strategy in **MSE**. The truth-telling equilibrium strategy is defined by the truthful report and first and second order beliefs are truthful. Our main interest in *Hypothesis 1c* is whether the measures of reciprocity significantly affect the likelihood that subjects play the truth-telling strategy.

The theory predicts that the truth-telling strategy is optimal for a moderate intensity of reciprocity which Figure 3.1 displays. We would therefore expect to observe that subjects would play the truth-telling strategy if the reciprocal measures is moderate. For our expected parameters this means that we would expect a significant positive linear effect of reciprocity and a significant negative quadratic effect of the reciprocal measure.

For both unincentivized measures we conduct there exists no significant relationship between reciprocity and the likelihood to play the truth-telling strategy. Independent of the used control the result holds. We conclude therefore that reciprocity cannot be accounted for the observed use of truth-telling strategy as predicted by the theoretical section.

Note that we observe that the frequency of agents that indeed play the truth-telling strategies is only around 10%, or in absolute numbers: only four subjects behave as if they are in the truth-telling BNFE. The number of observations is small because the sample size was calculated for treatment effects and the questions are not incentivized. These circumstances might explain the missing significant effect (Error Type II).

Beliefs about Roles' Behavior and Beliefs

In the analysis so far, we account at least partially the decrease in beliefs that the other subject reports truthful for the lower likelihood of implementation in the indirect mechanism. The framing of subjects into roles and their random allocation allows us to draw further conclusion for the decrease in beliefs.

Indeed, we find different treatments effects on the beliefs for the two roles which is illustrated in Figure 3.4. There, we observe that the first order beliefs of the seller do not differ between **BASE** and **MSE** ($p = 0.54$, Mann Whitney U Test). In contrast, the belief that the seller reports truthful by the buyers decreases significantly in **MSE** ($p < 0.01$, Mann Whitney U Test).

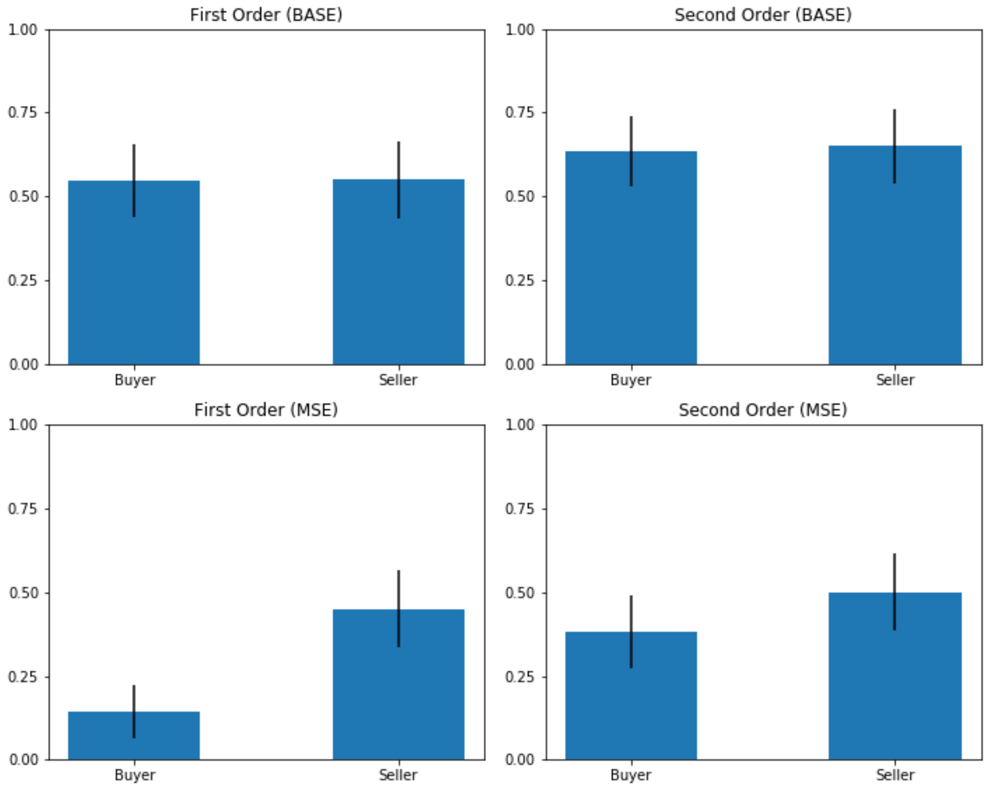


Figure 3.8: Difference in beliefs that the other subject truthfully reveals

This observation indicates that certain behavior, e.g. choosing kind options, is expected by some roles in the society but not by others. In our environment a seller's role might be associated with the objective to maximize profits (payoffs). This specific objective might induce a higher acceptability of selfish behavior, and hence the likelihood increases to use an unkind option to achieve it.

On the other hand sellers' second order beliefs reveal that they do not anticipate buyers' belief about sellers' behavior. On average buyers' assessments of sellers' behavior is correct as indeed only ten percent reveal their private information type. That also only ten percent of buyers are revealing is not anticipated by sellers. They seem to be to overoptimistic about the buyers' behavior and beliefs.

More general, the framing of a bilateral trade environment might be leading to the unsuccessful implementation of the social choice function by the indirect mechanism. Other environments, e.g. optimal provision of a public good, might lead to different beliefs about kind behavior and consequences among agents.

Robust Mechanism Design

In our third treatment, the Price Modification (**PM**), we construct the game induced by a modified minimal subsidy social choice function (Myerson and Satterthwaite, 1983). Although there is the disadvantage of a necessary subsidy by the experimenter to ensure voluntary participation, with respect to incentive compatibility this mechanism satisfies a robustness condition against arbitrary intensity of reciprocal preferences. Theoretically, in the induced game the truth-telling strategy is neither perceived as kind nor unkind. Therefore, the truth-telling strategy profile forms an equilibrium, either a BNE or a BNFE, for any kind of reciprocal concerns (including none). Indeed, in comparison to the less robust equilibrium in **MSE**, we find that

Result 3.3. *The likelihood that the truth-telling strategy is played is higher in **PM** than in **MSE**.*

Again, we find the result for both specifications of truthful revelation, i.e. only the observed behavior or conditional on the beliefs as the equilibrium strategy ($p < 0.01$, Mann Whitney U Test).

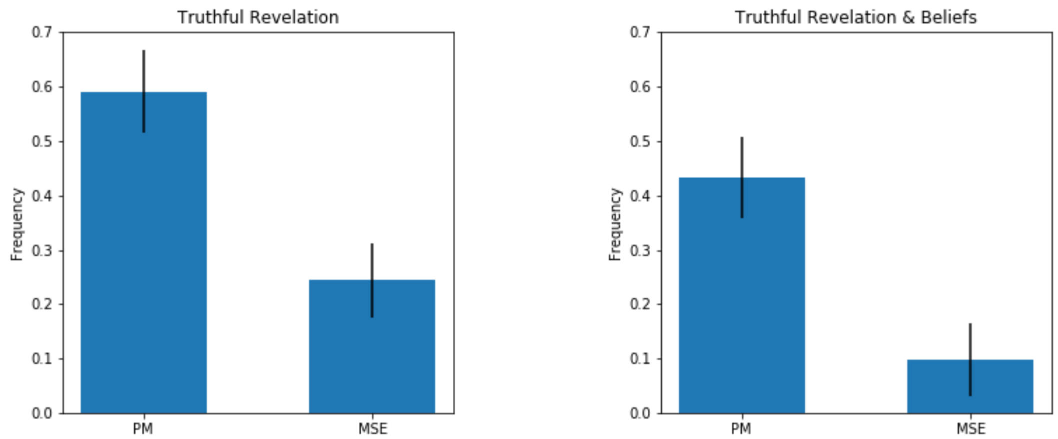


Figure 3.9: Difference in truthful revelation of the private information

In contrast, we do not find a significant increase in the truth-telling rate in **PM** in comparison to **BASE** with respect to pure observations ($p = 0.20$, Mann Whitney U Test) and conditional on beliefs ($p = 0.49$, Mann Whitney U Test).

As explained before there are multiple equilibria that lead to an efficient allocation in **MSE**. In contrast, in **PM** there exists only one equilibrium that ensures efficient trade. Hence, we test whether **PM** is more likely to establish an efficient allocation than **MSE**. Figure 3.10 shows the differences in the efficient allocations. As in our comparison before with **BASE**, we do not find any difference in the amount of efficient allocations between **MSE** and **PM** ($p = 0.67$, Mann Whitney U Test).

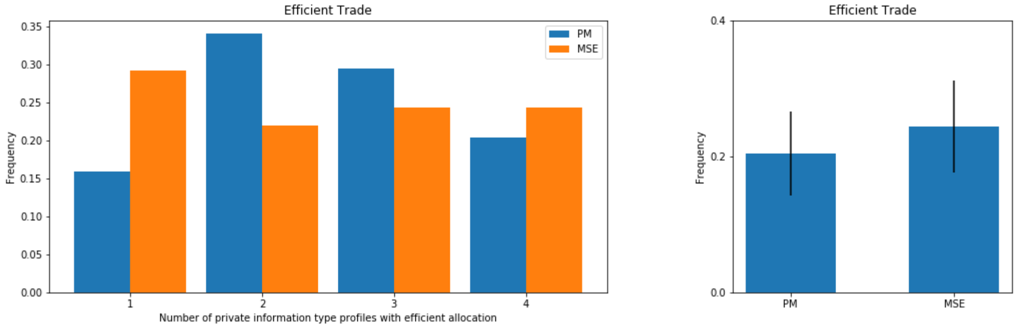


Figure 3.10: Difference in efficient allocations

Behavioral Welfare Economics

Another advantage of controlled laboratory experiments is the possibility to match survey questions with observed behavior. After the main experiment, subjects answer three questions about their happiness. In detail, the questions are about their current mood, whether they belief they are a 'good human' and their general life satisfaction.

Concerning behavioral welfare economics an increase in the happiness might be an indicator for an increase in welfare in a mechanism. In our case this is of special interest as the three mechanism do not significantly differ with respect to efficient allocations. In other cases, one could argue that there exists a trade-off between material gains and psychological happiness.

In a first step, we compare happiness differences for all three indicator for subjects if they play the truth-telling strategy. In **PM** the beliefs are such that the other subjects neither behaves kind nor unkind, while in **MSE** the subject beliefs that the other subjects behave kind. In the later the reaction of the believed kind behavior is kindness.

The *Hypothesis 4.1* states that this bilateral kindness increases the happiness (at least in mood) of the subjects. But we cannot reject the null-hypothesis that there are no differences between treatments. With Figure 3.11 we illustrate the average happiness of the different questions in **MSE** and **PM**.

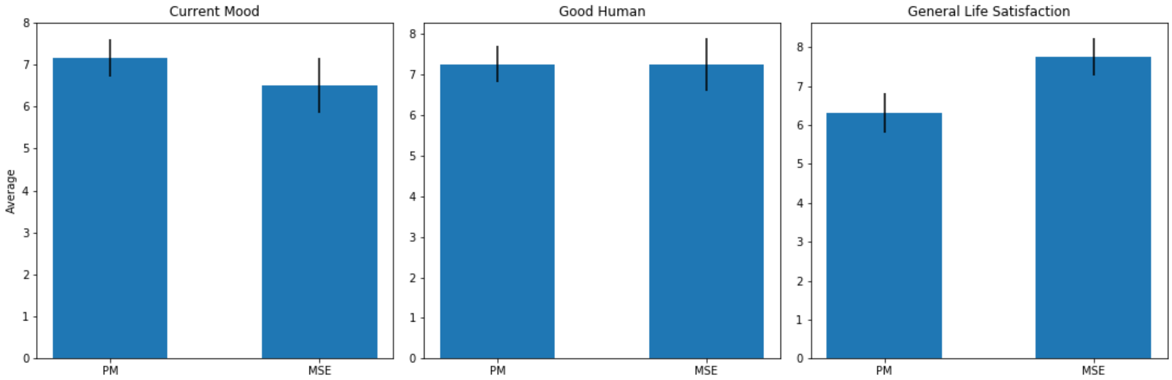


Figure 3.11: Difference in happiness in the truth-telling strategy

One potential problem could have been that the social choice function differs between the two treatments. Such a procedural fairness that the truthful revelation of private information yields an equal share of trade gains in comparison to one where this is ex post not more the case might lead already to different happiness measurements. As we depict in Figure 3.12 there is no general difference with respect to happiness between the two social choice functions considered in this Chapter.

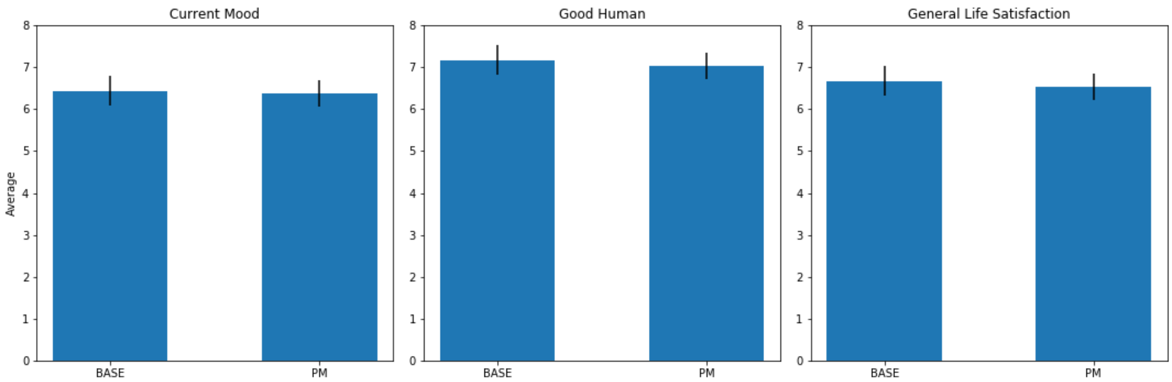


Figure 3.12: Difference in happiness in BASE and PM

Result 3.4. *Overall, we conclude that there are no systematic effects of the social choice function and the equilibrium concept on the happiness of experimental subjects.*

3.5 Discussion

In this chapter we test empirically several theoretical insights provided by Bierbrauer and Netzer (2016) in a controlled laboratory experiment. This empirical method enables us to control the environment and the random assignment of subjects to different roles and mechanisms. The experimental design matches most assumptions made in the underlying theory but leaves the behavior and beliefs to subjects.

In the first part of the analysis we do not observe empirical support for the theoretical counterexample to the revelation principle. Although the perceived kindness of the truth-telling strategy increase if unkind options are available, we find simultaneously a decrease in the beliefs that the other subject indeed plays the truth-telling strategy. This decrease is mainly driven by buyers' beliefs about sellers' behavior. The framing indicates that an empirical counterexample might be more likely to be observed in environments where kind behavior is rather expected.

An alternative experimental test of the theoretical prediction would have been to induce beliefs to subjects. Instead of unconditional behavior, one could condition subjects decision making on each option in the other role's message set. In other word, subjects report their fbest response to each possible strategy. The public good game literature finds that there exists a substantial number of subjects of so called conditional cooperation (Fischbacher et al., 2001). In these experiments the own strategy is conditioned on the strategy of other subjects. Here, we could allow subjects in our experiment that they report their strategy for each of the possible strategies of the other subject. We explicitly decided against this method because we believe that beliefs are crucial and worth studying for these environments. Also, conditional strategies would have complicated the experiment even further.

Although subjects build their beliefs about the other subject's behavior and beliefs freely, we hand out the complete instructions before the experiment. Hence, subjects know the questions concerning first and second order beliefs. This knowledge could induce strategic considerations which would have not exist without the questions. A potential problem here is that this strategic consideration could affect the formation of the beliefs and prime the relationship between behavior and beliefs. We accept this potential downside in our design because we want to ensure that subjects are aware of all payoff relevant factors. Otherwise we would not be able to control beliefs about the second part in the laboratory and its effect on the first part. We prefer to fix these beliefs to ensure that our treatment effects are robust against arbitrary beliefs in this dimension.

The equilibrium concepts considered here are constructed for a stage game of an one-time interaction between buyers and sellers. In the experiment subjects therefore play the game without repetition. This prohibits learning and thus updating of prior beliefs about potential behavior and beliefs of the assigned subject. On top the experiment is in a neutral frame without stating the properties of the good traded. Hence the belief formation is based on a very general consideration how buyer and seller in general (should) behave.

In the experiment we test the performance of different mechanisms. The idea is to modify the framework condition to achieve an efficient allocation. A free formation of beliefs is crucial because different framework modification might induce different beliefs about the assigned subject's behavior and beliefs. For instance, we observe that the possibility to be unkind decreases the belief that the other subjects truthfully reveal. Hence the free formation of beliefs allows us to observe the effect of different mechanisms on them.

For our horse race comparison of mechanisms we use an optimal mechanism for known reciprocal preferences (**MSE**) by Bierbrauer and Netzer (2016) as a reference. Based on pure selfish preferences an optimal mechanism (**PM**) is provided by Myerson and Satterthwaite (1983). In addition, the truth-telling strategies form an equilibrium for any intensity of reciprocal preferences in the game induced by this mechanism. Indeed, we find that the truth-telling strategy is more likely to be played in the more robust mechanism. Hence, the experiment also contributes to the discussion of robustness of mechanisms with respect to behavioral effects. The criticism of robust mechanism design is that optimal mechanisms often rely on specific knowledge about agents' beliefs and preferences. For most application this knowledge is not acquired by the designer, at least on an individual level. Many empirical studies report heterogeneity with respect to such parameters within the population.

Although the truth-telling strategies is more likely to be played in the **PM**, there are no differences to the **MSE** with respect to efficient allocations. In that sense the question remains whether it is important to control behavior or to solely focus on the implementation of efficient allocations. If we focus on the later we should note that the voluntary participation in **PM** is only guaranteed due to a subsidy by the experimenter and that there are no differences in the voluntary participation rate. Given the same efficiency and participating rate of the two mechanisms, one could argue that the **PM** is less desirable than the **MSE** because of the additional material costs for the designer. We also find that the **PM** does not increase significantly the rate that subjects truthfully reveal their private information in comparison to the direct mechanism tested in **BASE**. But with our experimental design we are not able to eliminate other explanations for deviating behavior as predicted by the standard theory. Alone in the domain of social preferences deviation might be purely due to outcome-based social preferences (Fehr and Schmidt, 1999; Saito, 2013).

The last two hypotheses might decrease the desirability of **PM** even further because we test whether subjects are less happy after participating in this mechanism.

These comparisons could contribute to the behavioral welfare economic literature. In classical terms, we define an allocation as efficient (welfare maximizing) if there is trade if and only if the valuation exceeds cost. This concept is based on material payoffs of agents. The behavioral literature finds that subjective well-being does not only focus on the own material well-being, but also include for instance procedural fairness and material well-being of the peer group.

Our experimental design allows for two different comparison of subjective well-being measures. First, we compare subjective well-being in two equilibria, once based on believed bilateral kindness and once not. Second, we compare the general subjective well-being in direct mechanisms that implement different social choice functions. While one explicitly shares gains of trade ex post between the two agents, the other one satisfies only ex ante equality.

For both cases we find no differences for all three subjective well-being measure. We used measurement from the literature concerning three categories: what is the current mood (1) and the general life satisfaction (2) and whether one considered oneself as a good human (3). The question is whether the formulation of the questions are most useful.

We took these three questions from the behavioral welfare literature to ensure credibility and comparability. Therefore, the question arises whether more specific questions for our environment would lead to the same observations. Another potential problem might be that in contrary to our predictions that the intention measurements are not significantly able to explain the truth-telling strategy and, hence, no observed difference in the subjective well-being measures.

3.6 Conclusion

In this chapter we test experimentally efficiency gains of bilateral trade mechanism based on believed intentions among agents. The analysis is based on Bierbrauer and Netzer (2016) who introduce the idea of social robustness in the mechanism design literature and provide a counterexample to the revelation principle and the Myerson-Satterwaithe impossibility theorem.

In our experiment we do not find empirical support for this counterexample. Our main explanation for the finding is the decreased beliefs in an bilateral environment that sellers are indeed kind if there is a monetary beneficial strategy available. A mechanism by Myerson and Satterthwaite (1983) that satisfies social robustness with respect to reciprocity provides another solution to the problem.

Although the social choice function is more likely to be implemented there is no difference in efficient allocation. With respect to voluntary participation and self-reported happiness we do not find any effect of the different mechanisms. As the robust solution relies on a subsidy from the outside, our findings indicate that the behavioral efficient mechanism might be more desirable.

3.A Equilibrium Analyses of the Games

In this section we provide a complete equilibrium analysis of the three games induced by the mechanisms introduced above. For each game the environment is such that that each combination of costs and valuation, θ, materializes with the same probability, i.e. $p(\theta) = 0.25$ for each $\theta \in \Theta$. We report the strategy profile s^* that forms either a Bayes Nash equilibrium (BNE) or a Bayes Nash Fairness equilibrium (BNFE) but implicitly assume each time that the first (and second) order beliefs are correct.

Our analysis start with the Bayesian game induced by the direct mechanism of the social choice function f^e. Table 3.5 displays the four complete information games that form joined with the probability distribution p the Bayesian game.

$(\underline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$10, 10$	$0, 0$
$\overline{\theta}_b$	$-30, 50$	$-70, 90$

$(\underline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$10, -70$	$0, 0$
$\overline{\theta}_b$	$-30, -30$	$-70, 10$

$(\overline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$90, 10$	$0, 0$
$\overline{\theta}_b$	$50, 50$	$10, 90$

$(\overline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$90, -70$	$0, 0$
$\overline{\theta}_b$	$50, -30$	$10, 10$

Table 3.5: Game induced by an direct mechanism of the SCF f^e

There exist two BNE in the Bayesian game: (a) $((\underline{\theta}_b, \underline{\theta}_b), (\underline{\theta}_s, \overline{\theta}_s))$ or (b) $((\underline{\theta}_b, \overline{\theta}_b), (\overline{\theta}_s, \overline{\theta}_s))$. In none of them both agents report truthful their type. Both of them remain BNFE if the reciprocity weighting $\gamma < 0.02$ because their payoff advantage still outweighs the possibility to reciprocate unkind behavior by reporting always the weak type. Indeed, there exists for any $\gamma > 0$ the (unkind) BNFE $((\underline{\theta}_b, \underline{\theta}_b), (\overline{\theta}_s, \overline{\theta}_s))$ where both agents always report the weak type and hence trade never takes place. Given $\gamma \geq 0.05$ the strategy profile $((\overline{\theta}_b, \overline{\theta}_b), (\underline{\theta}_s, \underline{\theta}_s))$ forms a (kind) BNFE. To conclude, none of the equilibria implements the social choice functions.

As the direct mechanism does not implement we expand the message sets of both agents. The buyer is able to understate the low valuation ($\tilde{\theta}_b$) and the seller to overstate the high cost ($\tilde{\theta}_s$). The consequences are as if they would have stated the low (or high respectively) type but receive additional composition of $\delta_b = \delta_s = 5$. In Table 3.6 we display the four complete games that form with the probability mass function $p(\cdot)$ the Bayesian game by such an mechanism.

$(\underline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$	$\tilde{\theta}_s$
$\tilde{\theta}_b$	$10+\delta_b, 10-\delta_b$	$0, 0$	$0, 0$
$\underline{\theta}_b$	$10, 10$	$0, 0$	$0, 0$
$\overline{\theta}_b$	$-30, 50$	$-70, 90$	$-70-\delta_s, 90+\delta_s$

$(\underline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$	$\tilde{\theta}_s$
$\tilde{\theta}_b$	$10+\delta_b, -70-\delta_b$	$0, 0$	$0, 0$
$\underline{\theta}_b$	$10, -70$	$0, 0$	$0, 0$
$\overline{\theta}_b$	$-30, -30$	$-70, 10$	$-70-\delta_s, 10+\delta_s$

$(\overline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$	$\tilde{\theta}_s$
$\tilde{\theta}_b$	$90+\delta_b, 10-\delta_b$	$0, 0$	$0, 0$
$\underline{\theta}_b$	$90, 10$	$0, 0$	$0, 0$
$\overline{\theta}_b$	$50, 50$	$10, 70$	$10-\delta_s, 90+\delta_s$

$(\overline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$	$\tilde{\theta}_s$
$\tilde{\theta}_b$	$90+\delta_b, -70-\delta_b$	$0, 0$	$0, 0$
$\underline{\theta}_b$	$90, -70$	$0, 0$	$0, 0$
$\overline{\theta}_b$	$50, -30$	$10, 10$	$10-\delta_s, 10+\delta_s$

Table 3.6: Game induced by the indirect mechanism Γ^e_{ind}

The set of BNE is similar to the one of the game induced by the direct mechanism. In contrast, as the weak type is weakly dominated by the new action there exist the two BNE $(\tilde{\theta}_b, \tilde{\theta}_b), (\underline{\theta}_s, \tilde{\theta}_s))$ and $((\tilde{\theta}_b, \overline{\theta}_b), (\tilde{\theta}_s, \tilde{\theta}_s))$. But since both agent are indifferent between their weak types and the new actions if no trade takes place, $((\tilde{\theta}_b, \tilde{\theta}_b), (\underline{\theta}_s, \overline{\theta}_s))$ and $((\underline{\theta}_b, \overline{\theta}_b), (\tilde{\theta}_s, \tilde{\theta}_s))$ form to additional BNE who are payoff equivalent to the first two.

The parameterization is chosen such that with a sufficient modest weighting for reciprocity, $0.2 \leq \gamma \leq 0.45$, the truth-telling strategy $s^{TT} = ((\underline{\theta}_b, \overline{\theta}_b), (\underline{\theta}_s, \overline{\theta}_s))$ forms indeed a BNFE. In addition, for any possible $\gamma > 0$ there exists the (kind) BNFEs $((\overline{\theta}_b, \overline{\theta}_b), (\underline{\theta}_s, \underline{\theta}_s))$ where inefficiently much trade takes place. On the opposite there exists a set of (unkind) BNFE with no trade. Given that there is concern for reciprocity for any $\gamma > 0$ any combination where the buyer uses either $(\underline{\theta}_b, \underline{\theta}_b)$, $(\underline{\theta}_b, \tilde{\theta}_b)$, $(\tilde{\theta}_b, \underline{\theta}_b)$ or $(\tilde{\theta}_b, \tilde{\theta}_b)$ and the seller $(\overline{\theta}_s, \overline{\theta}_s)$, $(\overline{\theta}_s, \tilde{\theta}_s)$, $(\overline{\theta}_s, \tilde{\theta}_s)$ or $(\tilde{\theta}_s, \tilde{\theta}_s)$ forms such an (unkind) BNFE.

The last part of equilibrium analysis concentrates on the game induced by the minimal subsidy social choice functions. As before, in Table 3.7 we display the four complete information games.

$(\underline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$0, 30$	$5, 5$
$\overline{\theta}_b$	$-25, 55$	$-50, 80$

$(\underline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$0, -50$	$5, 5$
$\overline{\theta}_b$	$-25, -25$	$-50, 0$

$(\overline{\theta}_b, \underline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$80, 30$	$5, 5$
$\overline{\theta}_b$	$55, 55$	$30, 80$

$(\overline{\theta}_b, \overline{\theta}_s)$	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$80, -50$	$5, 5$
$\overline{\theta}_b$	$55, -25$	$30, 0$

Table 3.7: Game induced by an direct mechanism of the SCF f^*

The game is designed such that the truth-telling strategy forms a BNE. Nevertheless, this equilibrium is not unique. If a agent beliefs the other agent reports truth-fully always report the weak type is additionally a best response. Hence there a two less efficient BNE such that either $((\underline{\theta}_b, \underline{\theta}_b), (\underline{\theta}_s, \overline{\theta}_s))$ or $((\underline{\theta}_b, \overline{\theta}_b), (\overline{\theta}_s, \overline{\theta}_s))$ is played in the game. Considering sufficiently small reciprocal concerns, i.e. $\gamma < 0.15$, both strategy profiles form also an BNFE. In addition, if $\gamma \geq 0.05$ there exists the (kind) BNFE $((\overline{\theta}_b, \overline{\theta}_b), (\underline{\theta}_s, \underline{\theta}_s))$ with inefficiently much trade and if $\gamma \geq 0.08$ the (unkind) BNFE $((\underline{\theta}_b, \underline{\theta}_b), (\overline{\theta}_s, \overline{\theta}_s))$ with no trade at all. But of most interest is that in the game for any $\gamma > 0$ the truth-telling strategies form also BNFE which is neither kind nor unkind. Hence, for any reciprocal weighting parameter the game induced by the direct mechanism implements the social choice function f^*.

3.B Equilibrium Figures

In the following we display the equilibrium play for each role given the first order beliefs about the other player's behavior. Note that this is different from best response play. The requirement is that the player anticipates that the other player plays a best response to the own behavior, or in other words, we assume correct second order beliefs. A red circle indicates that this strategy profile forms a Bayes Nash equilibrium. A blue circle indicates that this strategy profile forms a Bayes Nash Fairness equilibrium for some interval of γ. The concrete parameters of γ such that a strategy profile forms a Bayes Nash Fairness equilibrium is provided in the previous Appendix Section 3.A.

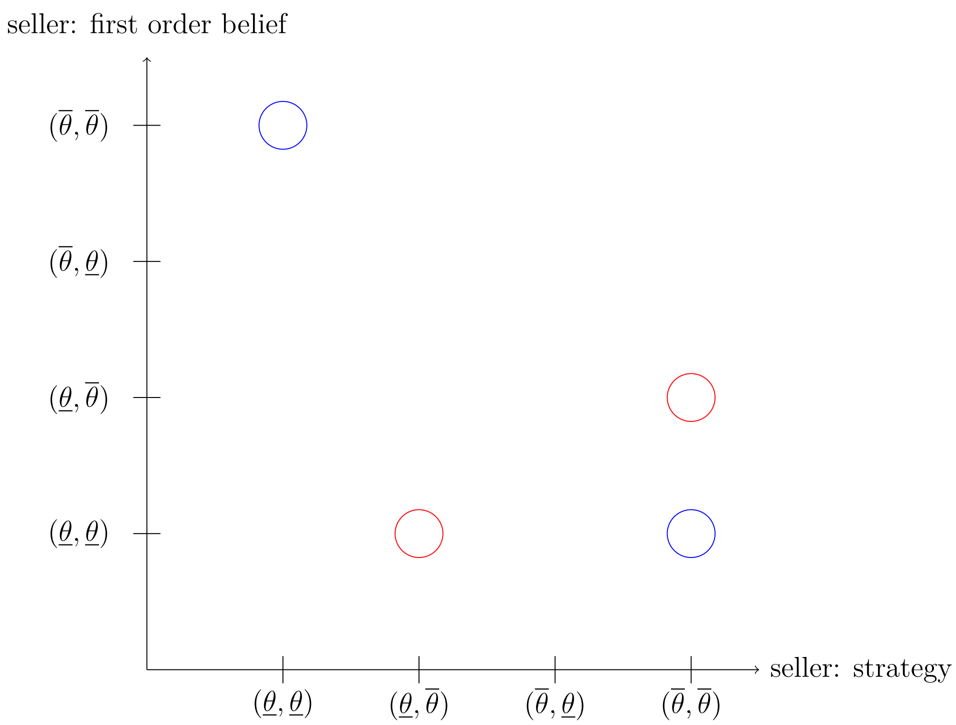


Figure 3.13: Equilibria in the BASE treatment: seller's point of view

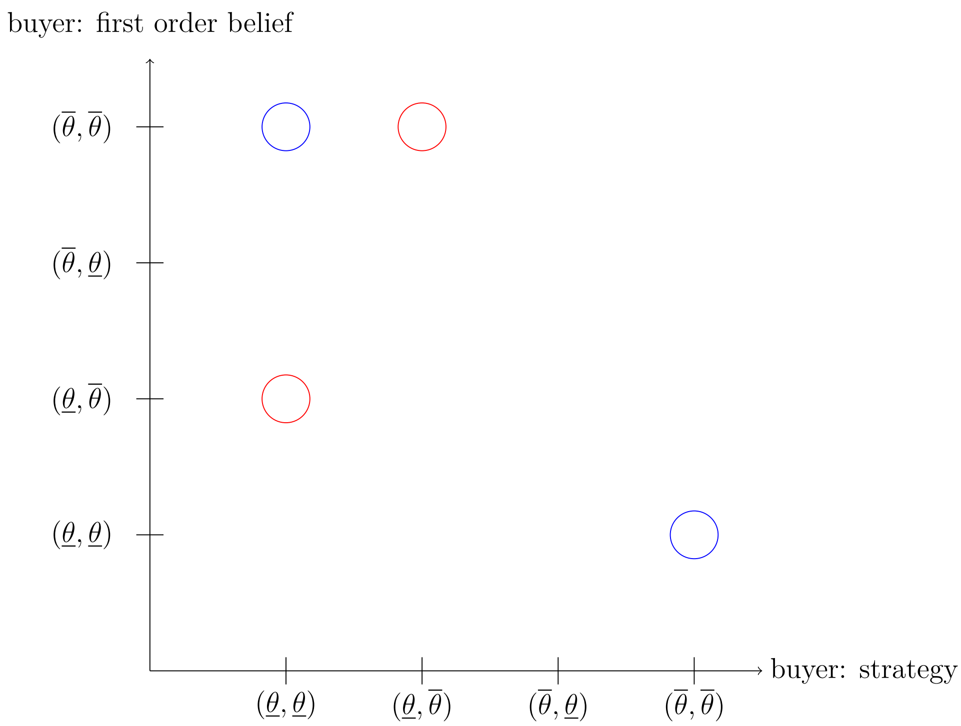


Figure 3.14: Equilibria in the BASE treatment: buyer's point of view

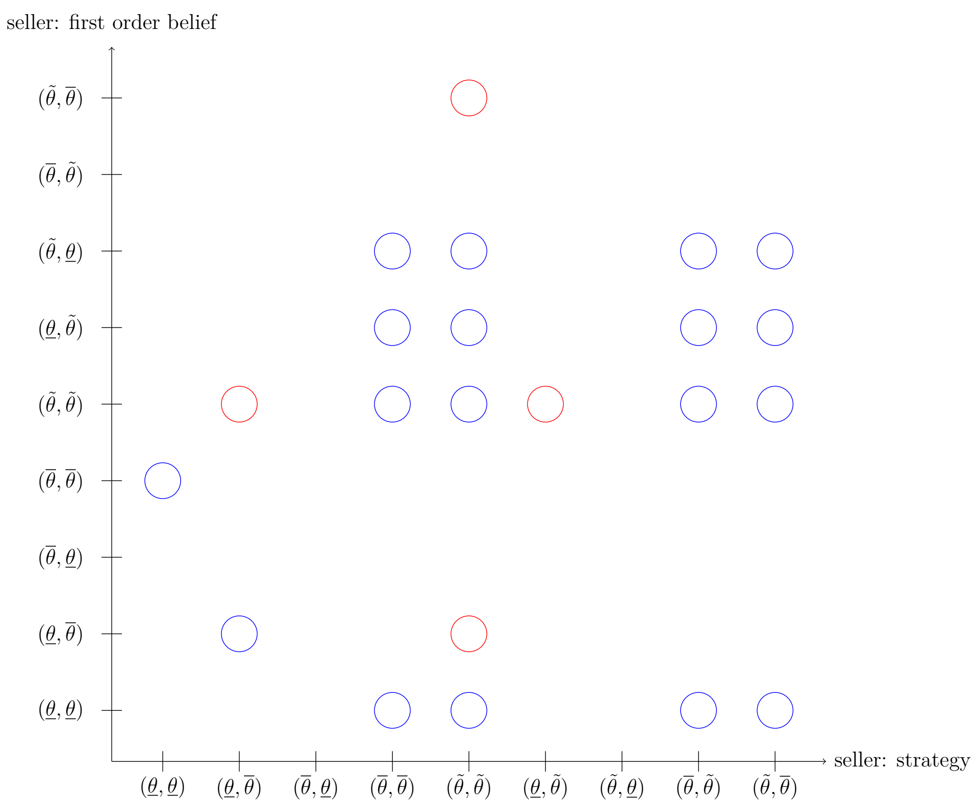


Figure 3.15: Equilibria in the MSE treatment: seller's point of view

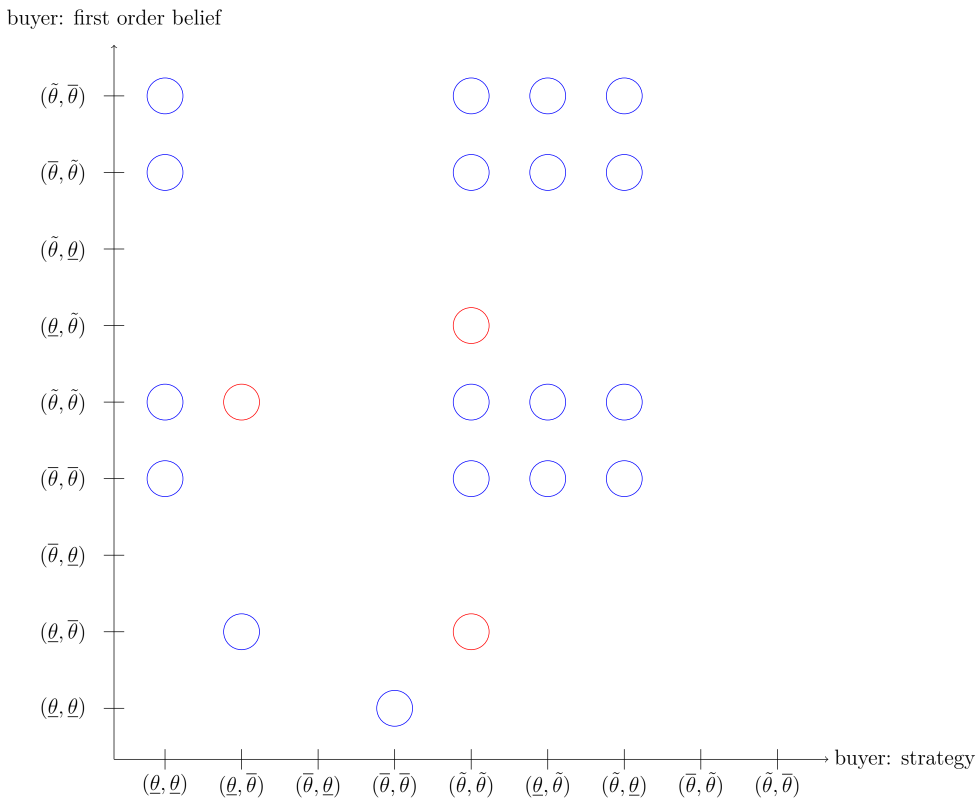


Figure 3.16: Equilibria in the MSE treatment: buyer's point of view

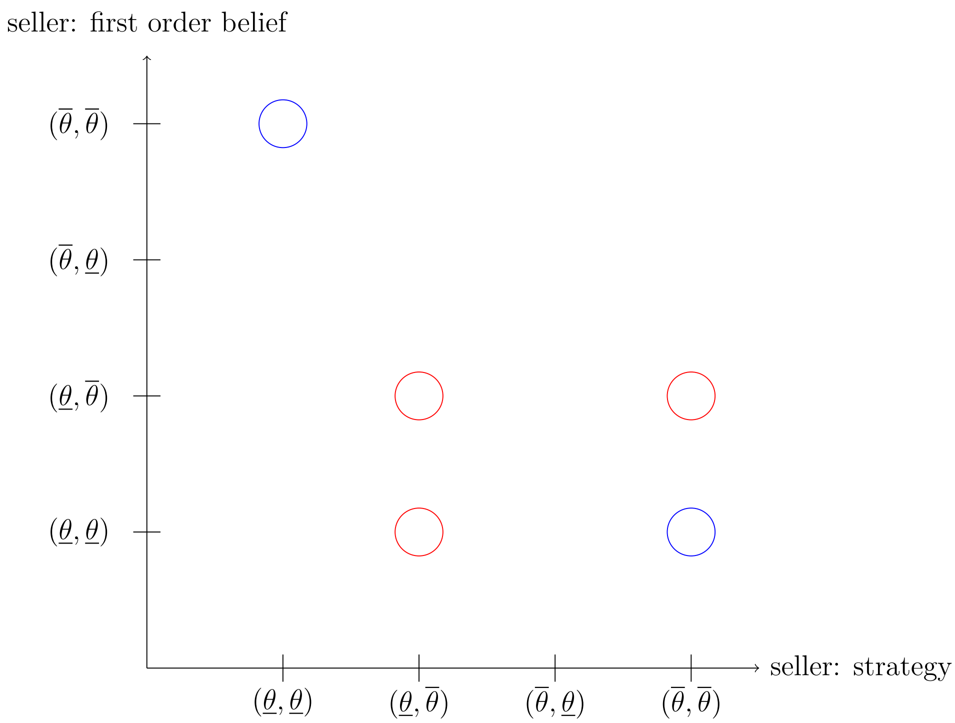


Figure 3.17: Equilibria in the PM treatment: seller's point of view

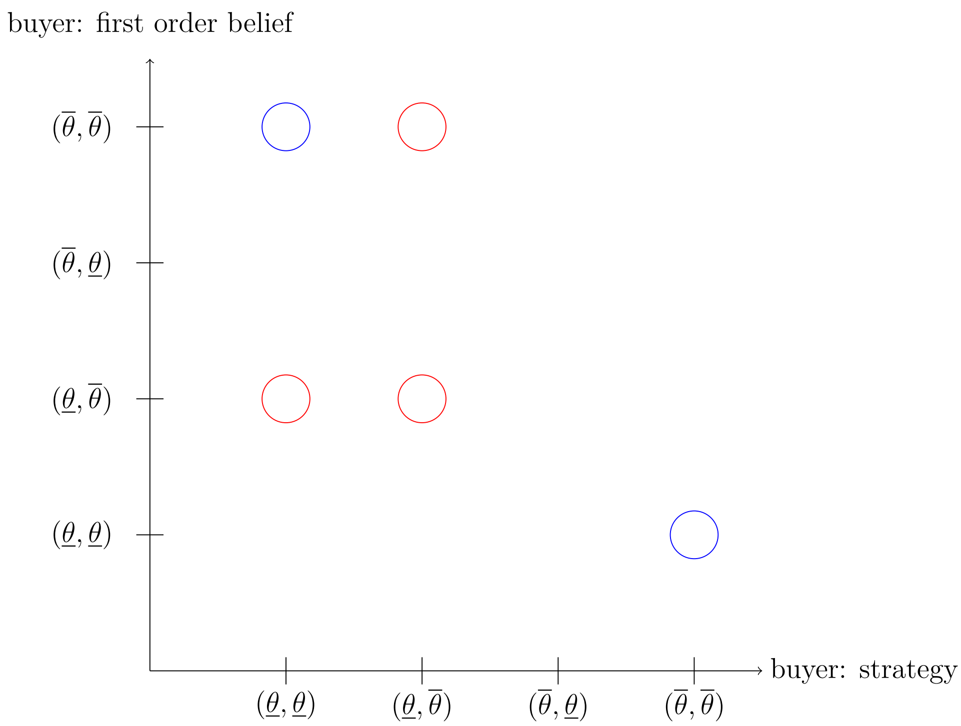


Figure 3.18: Equilibria in the PM treatment: buyer's point of view

3.C Subjects Behavior and Beliefs

In this section we state the observed behavior of subjects in all three treatments given their beliefs. While in the BASE treatment less than 50% of the observed behavior can be explained by equilibrium play ($\frac{6}{21}$ for buyers and $\frac{9}{21}$ for sellers), we find that in MSE (buyer: $\frac{12}{21}$, seller: $\frac{13}{21}$) and in PM (buyer: $\frac{17}{22}$, seller: $\frac{15}{22}$) the majority behaves as if they are in one of the equilibria. Note that we did not include all possible best response given their belief in this calculation. The notation in following figures is such that $\underline{\theta} = \theta_L$, $\overline{\theta} = \theta_H$ and $\tilde{\theta} = \theta_X$.

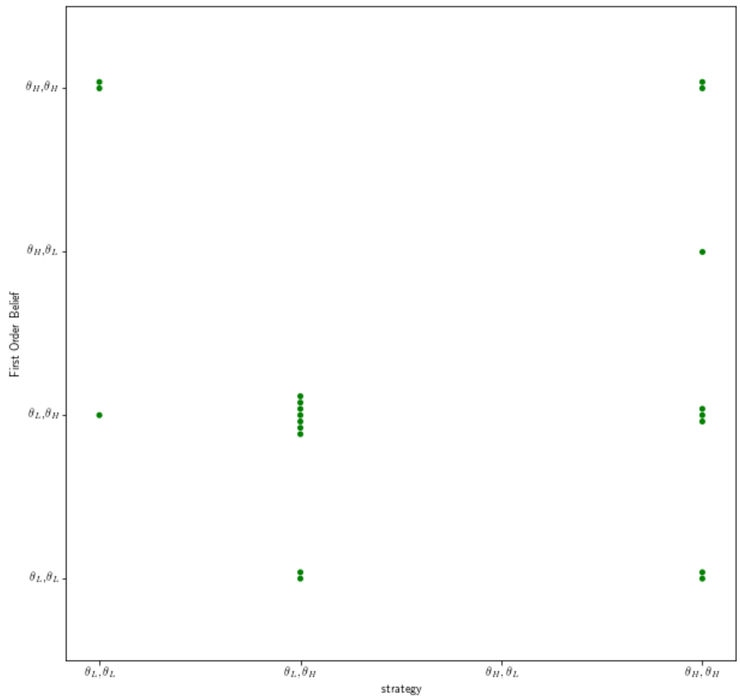


Figure 3.19: Sellers' behavior and corresponding beliefs in BASE

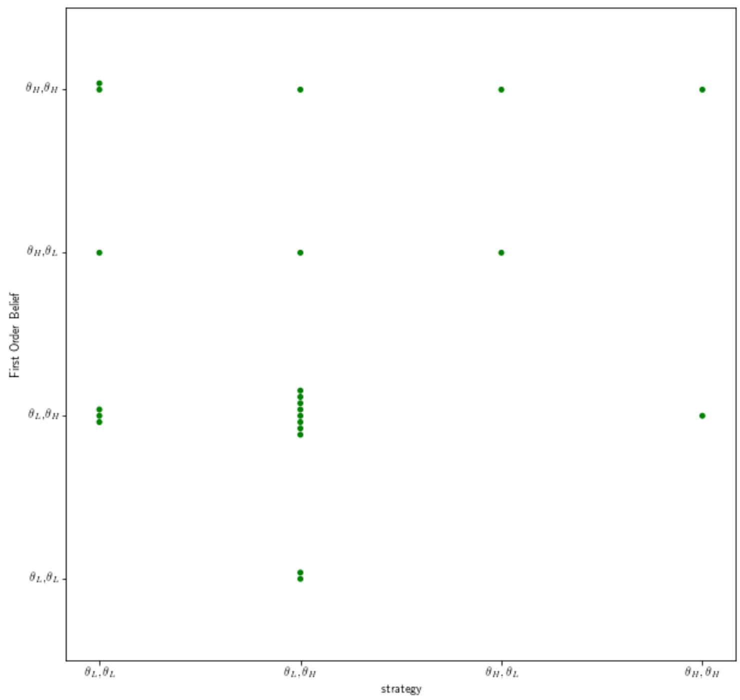


Figure 3.20: Buyers' behavior and corresponding beliefs in BASE

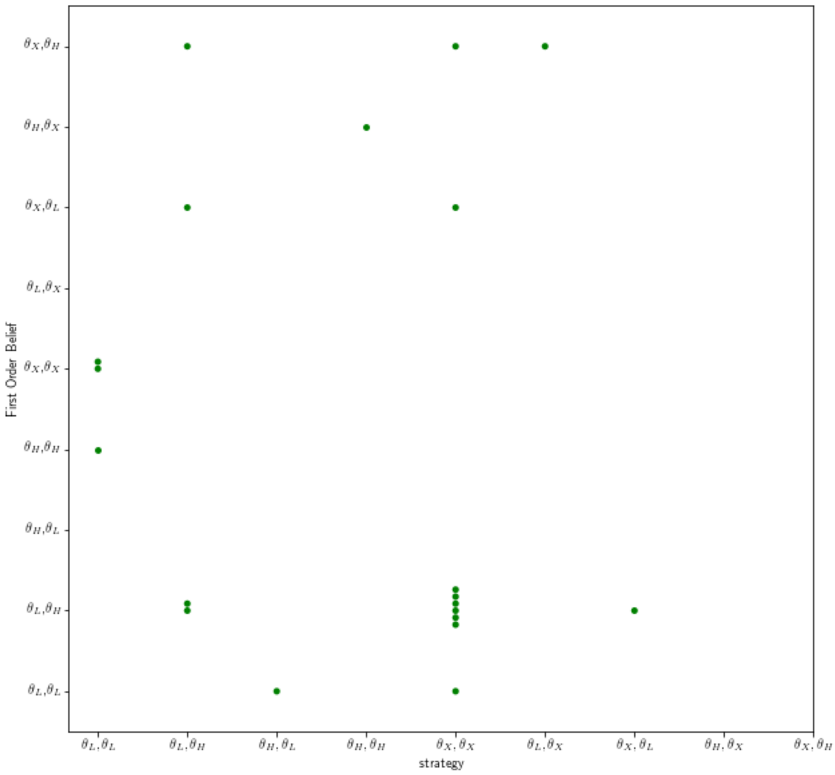


Figure 3.21: Sellers' behavior and corresponding beliefs in MSE

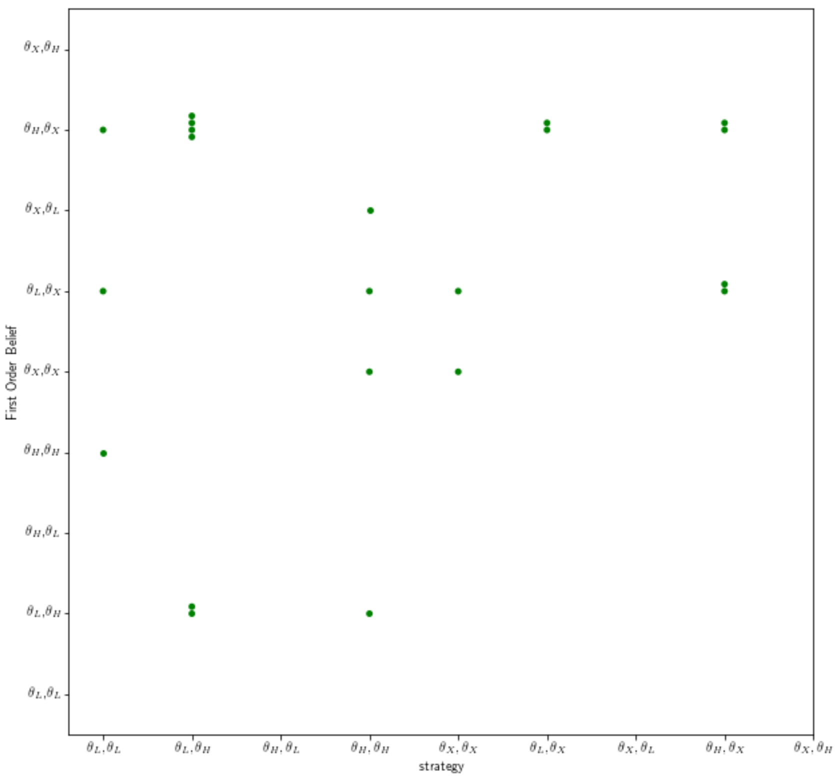


Figure 3.22: Buyers' behavior and corresponding beliefs in MSE

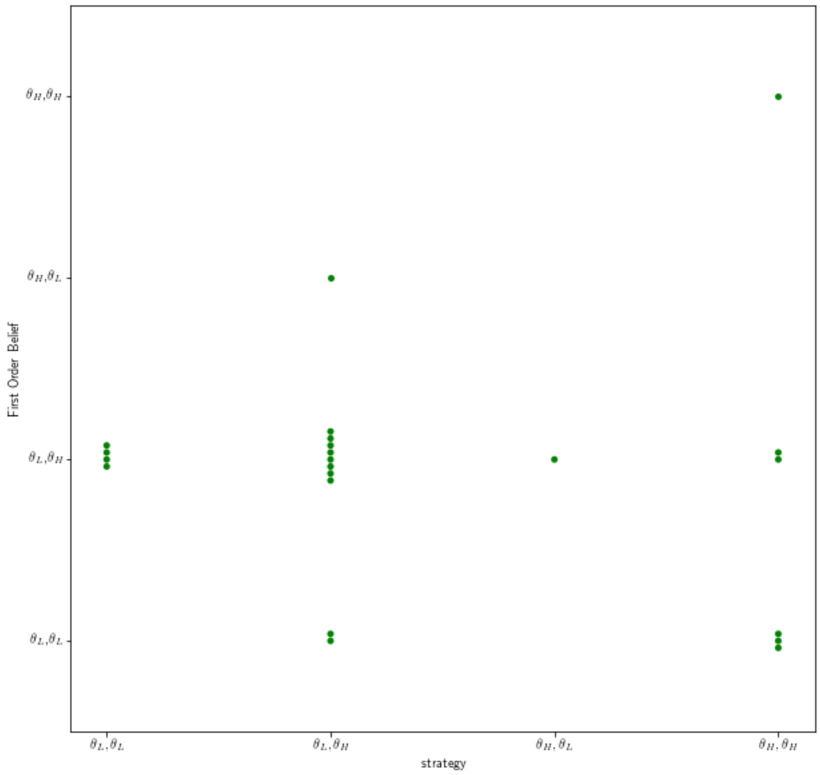


Figure 3.23: Sellers' behavior and corresponding beliefs in PM

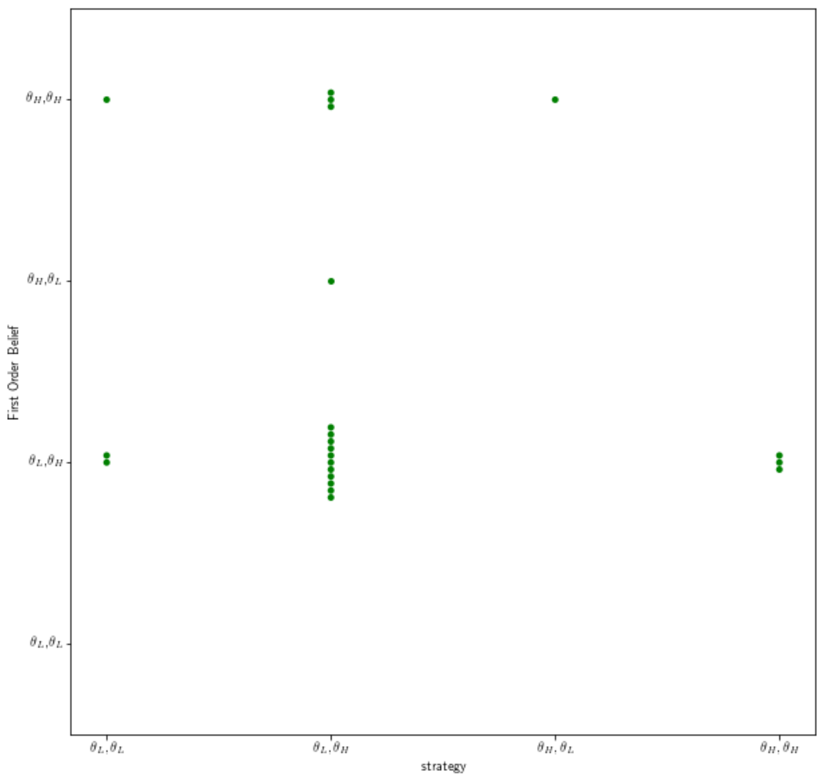


Figure 3.24: Buyers' behavior and corresponding beliefs in PM

3.D Voluntarily Participation

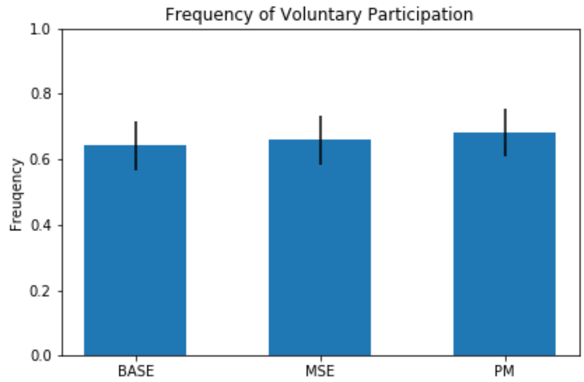


Figure 3.25: Voluntary participation by treatment

We find no difference in the rate of voluntarily participation between the treatments.

3.E Simulation for Subjects' Understanding

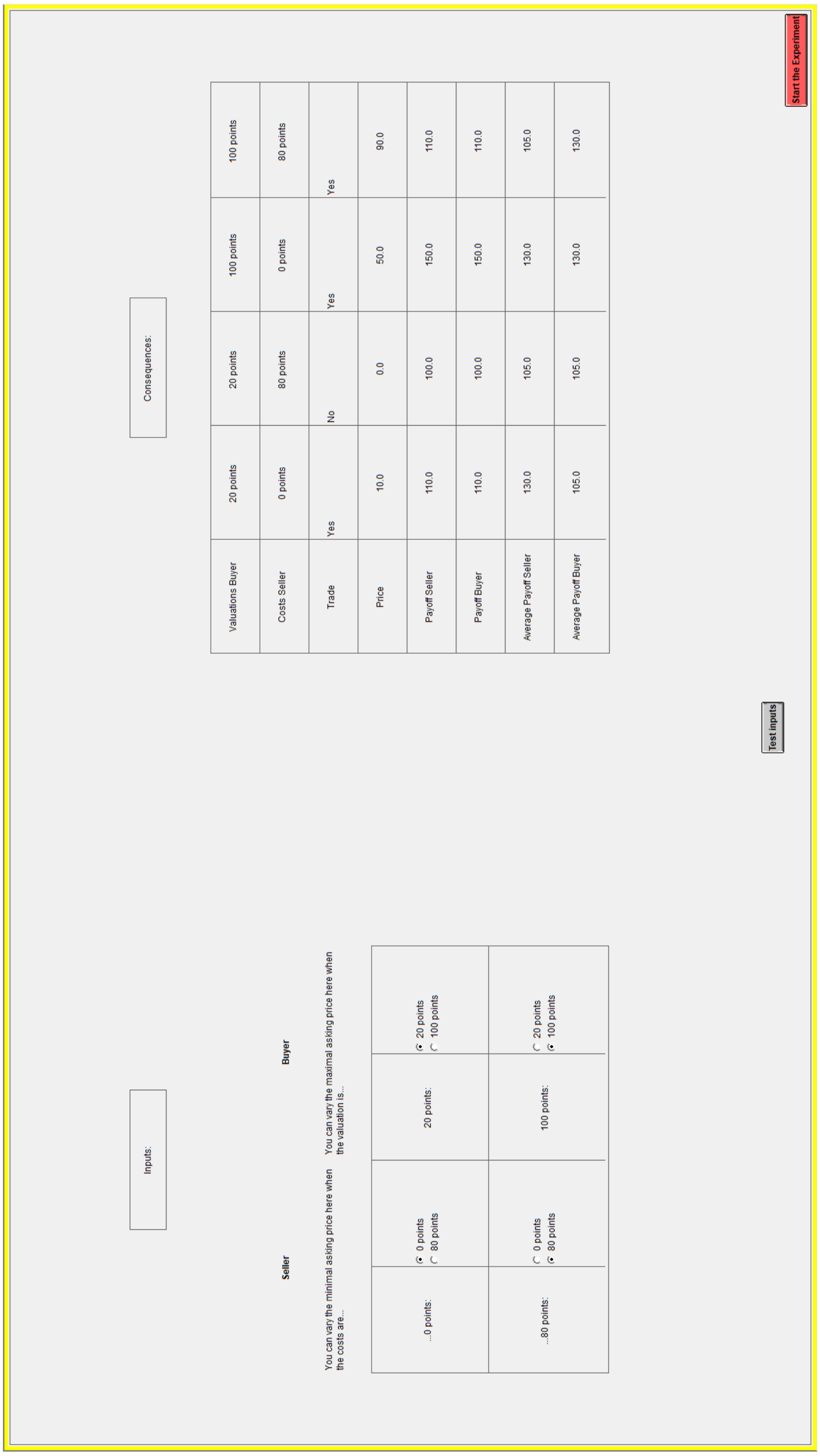


Figure 3.26: Simulation Screen in BASE for truth-telling strategies

To increase the understanding of the game subjects could before and during the first part simulate consequences for strategies of both roles. On the left side of the screen they are able to state a strategy for both roles. In this stage subjects are unaware of their own role later in the experiment. Hence, they learn the environment from a neutral perspective.

For each strategy combinations entered in the left side, all relevant information are provided on the right side of the screen. In the beginning no strategy is clicked to avoid any default or anchoring effects. The Figure 3.26 illustrates the situation where the subjects simulates the truth-telling strategy for both roles. When they click on the "Testing Inputs"-Button the consequences on the right side are updated.

3.F Questionnaire Detail

Online experiment

1) How willing are you to punish someone who behaved unfair towards **you**, even if it is costly for you?
2) How willing are you to punish someone who behaved unfair towards **others**, even if it is costly for you?
Likert Scale from 0 = "absolutely not willing" to 10 = "absolutely willing"

3) If someone does me a favor, I am willing to a favor, too.
4) If someone does unjust to me, I am willing to revenge myself in the first chance.
5) General I belief that other people have good intention.
Likert Scale from 0 = "Does not describe my character" to 10 = "Describes my character"

Please imagine the following situation:
You are in a for you unusual area and have lost your way. You ask an unknown for the way. This person offers you to bring you to your target. In order to help you, the person has cost of 20 Euro. But he is not accepting any monetary reward in return. Coincidentally you have six presents with you. The cheapest costs 5 euro, the most expensive one 30 Euro. Would you give to the person a "Thank you-"present? If so, which present would you choose?
No, 5, 10, 15, 20, 25, 30

Laboratory experiment

1) In general, I consider myself:
from 1 = "not a very happy person" to 7 = "a very happy person"

2) How satisfied are you currently, all in all, with your life?
from 0 = "completely unsatisfied" to 10 = "completely satisfied"
3) What is your mood at the moment?
from 0 = "very bad" to 10 = "very good"
4) When you would had the possibility to remain from the trade and keep your 100 points, would you have then participated in the trade?
1 = "Yes, I would have participated in the trade" and 0 = "No, I would not have participated in the trade."
5) We would like you to rate the four possibilities of the other subject. The first number of points is the minimal (maximal) asking price if the costs are 0 (the valuation is 20). The second number of points is the minimal (maximal) asking price if the costs are 20 (the valuation is 100).
0,0 or 20,20
0,80 or 20, 100
80,0 or 100, 20
80,80 or 100, 100
(in the **MSE** subjects also rate the five remaining strategies including the unkind message)
from 0 = "unkind" to 10 = "kind"
6) We would like you to report which decision you would have reported, if you would have been in the role of the other subject.
a) Which minimal (maximal) asking price would you have reported, if your costs are 0 (your valuation is 20) points:
Buyer: 0, 80, [90]; Seller: [10], 20, 80
b) Which minimal (maximal) asking price would you have reported, if your costs are 80 (your valuation is 100) points:
Buyer: 0, 80, [90]; Seller: [10], 20, 80

3.G Experimental Instructions

In the following we display the instructions for the three treatments. The instructions are presented from the Baseline (BASE) treatment. The modifications for the Message Set Expansion (MSE) treatment are in red, while the modifications for the Price Modification (PM) treatment are in green.

Instructions

Welcome to our experiment! When you carefully read the following instruction, you can - depending on your decisions - earn a non-inconsiderable amount of money. So, it is important, that you read carefully this instruction. During the experiment there exists an absolute prohibition of communications among experimental participants. The neglect of this rule will lead to the exclusion of the experiment and the payment. When you have questions, please raise your hand. We will come to you in that case. During the experiment we do not talk about euros, but rather of points. Your complete payment will be calculated first in points. The earned points in the experiment will be converted at the end of the experiment into euro with a rate of

$$20 \ points \ = \ 1 \ Euro.$$

At the end of the experiment today you will receive your earned points in cash. Either Part 1 or Part 2 of todays experiment is relevant for the payment. This is decided randomly with equal chances of both. Additionally, each participant receives a 5 Euro show up fee and the payment from the online part.

For today's experiment you will be randomly assigned with another participants. In both parts of the experiment you are assigned with the same participant.

Part 1

The both roles in the experiment

In this part, you and your assigned participants are assigned to one of two different roles. In the case you receive the role as Seller, the assigned participant receives the role as Buyer. In the case you receive the role as Buyer, the assigned participant receives the role as Seller. Both participants are endowed with a initial endowment of 100 points.

The Seller has the possibility to produce a good. The production of the good has costs of either 0 or 80 points. Both costs are equally likely (50%/50%). This means, if the Seller produces the good, the costs (0 or 80 points) will be deducted from the own initial endowment at the end of the experiment.

The Buyer has the possibility to buy the good and has a valuation for the good of either 20 or 100 points. Both valuations are equally likely (50%/50%). This means, if the Buyer buys the good, the valuations (20 or 100 points) will be added to the own initial endowment at the end of the experiment.

The workings of trade

The Seller does not know the size of the costs at the point in time when the decision has to be made. Therefore, the Seller is able to report a plan. This means, for both costs (0 and 80 points) the Seller has the opportunity to report a minimum asking price (0 or 80 or 90 points). In the end of the experiment only one of the two costs are payoff relevant. Therefore, also only the reported minimum asking price in that case is payoff relevant. At the point in time when the decision has to be made, the Seller does not know the valuation of the Buyer.

The Buyer does not know the size of the valuations at the point in time when the decisions has to be made. Therefore, the Buyer is able to report a plan. This means, for both valuations (20 and 100 points) the Buyer has the opportunity to report a maximum asking price (10 or 20 or 100 points). In the end of the experiment only one of the two valuations are payoff relevant. Therefore, also only the reported maximum asking price in that case is payoff relevant. At the point in time when the decision has to be made, the Buyer does not know the cost of the Seller.

The computer randomly chooses for both roles which cost (0 or 80 points) and valuation (20 or 100 points) is relevant. For both roles both possibilities are equally likely. For instance, is the probability that the cost of the Seller is equal to 0 points and the valuation of the Seller 20 points 25% (50%*50%).

The computer then compares the for the randomly chosen cost and valuation the stated minimum asking price with the stated maximum asking price.

If the maximum asking price of the Buyer is smaller than the minimum asking price of the Seller, then the good will not be produced. In this case, we say that there is no trade.

If the maximum asking price of the Buyer is larger than the minimum asking price of the Seller, then the good will be produced. In this case, we say that there is trade.

In the case there is trade, the Buyers hast to pay the price for the good to the Seller. The price is calculated in the following way.

$$Price = \frac{minimum\ asking\ price\ +\ maximum\ asking\ price}{2}$$

The price increases additionally by 15 points, if the minimal asking price is 0 points and the maximal asking price is 20 points.

The price decreases additionally by 15 points, if the minimal asking price is 80 points and the maximal asking price is 100 points.

In the following table you can see the calculated price for the different asking prices:

		minimum asking price		
	price	0	80	90
maximum asking price	0	5	(no trade)	(no trade)
	20	10 (25)	(no trade)	(no trade)
	100	50	90 (75)	95

Table 3.8: Price calculations

Independent whether there is trade or not, every participants receives additional 5 points.

In the case that there is no trade both roles keep their initial endowment of 100 points.

The following table summaries again, how the payments at the end are calculated:

payment	no trade	trade
Seller	initial endowment + 5	initial endowment cost + price + 5
Buyer	initial endowment + 5	initial endowment + valuations - price + 5

Table 3.9: Payment calculations

At the end of the experiment you will be informed about,

(a) what are the minimum and maximum asking prices,

(b) whether there is trade and if so, atwhich price, and

(c) your payments in points.

Before the real experiment starts, you are able to simulate all possible scenarios. These scenarios are not payoff relevant. You have the possibility to look at the payments for different costs, valuations as well as minimum and maximum asking price for both roles. During the first part of the experiment you are able to go back to this simulation.

Part 2

In the second part of the experiment we will ask you questions concerning the behavior of you and your assigned participant. For each correct answered payoff relevant question, you receive 30 points. In total, there will be four payoff relevant questions such that you can earn up to 120 points in this part. The question will be displayed on the screen after you finished the first part.

The first two questions are about your assessment about the behavior of the assigned participant. You will be asked which asking prices do you think has the assigned participants reported, when the costs/valuations are either low or high.

The last two questions are about your assessment about the answers that the assigned participant has reported about your behavior. You will be asked which assessment the assigned participants in your belief has reported, when your costs/valuations are either low or high.

Chapter 4

The Dependence of Crémer-McLean Auctions on Selfish Preferences

4.1 Introduction

For the last decades the economic literature acknowledges the importance of social motives in many information economic environments, e.g. public good provision and auctions. Experimental studies on optimal auction design show for instance that there is a tendency of overbidding in second price auctions (SPA) (Kagel et al., 1987). One potential explanation for this behavior is that bidders are spiteful among each other (Morgan et al., 2003; Andreoni et al., 2007). Bartling and Netzer (2016) modify the SPA into an external-robust auction (ERA) where potential influence on other bidder's payoffs is eliminated. The authors demonstrate in an experimental study that bidders' truthful revelation of private information is indeed more likely in the ERA than in the SPA. Their findings support the idea of social motives between bidders in auctions and its adverse effect on efficient allocations.

In general, the information economic literature observes that in environments with asymmetric information the privately informed party is granted an information rent. Exceptions of this insight are generic cases of correlated valuation types (Crémer and McLean, 1985, 1988). Arguably, for many environments this seems to be the more realistic assumption. The environment we consider here is an auction of a single unit good with two bidders.

In this paper, we study the robustness of the auction by Crémer and McLean (1985) towards outcome-based social preferences. We ask for which kind of social preferences the first best implementation in the sense of Crémer and McLean (1985) is possible. For the analysis we apply the utility function by Charness and Rabin (2002) which represents many kinds of preferences concerning differences to the other bidder's material well-being.

In the case of correlated valuation types the assumption is that bidders who differ in their valuation types, also differ in their beliefs about the other bidder's valuation type. For example, a bidder might be more likely to belief that the other bidder's valuation type is high, if she herself has a high valuation type. To be concrete, bidders' valuations are then mapped to a corresponding belief distributions about the other bidder's valuation types. The Crémer-McLean assumption states that these belief distributions need to be linear independent. In that case, the auction mechanism by Crémer and McLean (1985) exploits this difference in belief distributions for the full extraction of the expected surplus generated by the trade.

The literature on correlated valuation types expands this basic result. For instance, McAfee and Reny (1992) shows that the full surplus extraction is also possible in the case of continuous type sets. More recently, Gizatulina and Hellwig (2017) prove that the result holds for a generic set of private information types. Kosenok and Severinov (2008) show that in combination with an additional constraint the Myerson-Satterthwaite Impossibility Theorem (Myerson and Satterthwaite, 1983) no longer applies if valuation types are correlated. Hence, in such environments the implementation of social choice functions that satisfy individual rationality, Bayesian incentive compatibility and ex post budget balance is possible.

But the applicability of first best implementation has been criticized. For example, Robert (1991) points out that the result no longer holds, if either risk neutrality or unlimited liability is violated (which is already mentioned by Crémer and McLean, 1985, themself). Neeman (2004) criticizes that the result relies on the assumption that the preferences are uniquely determined by the own information type and shows that this assumption is necessary for full surplus extraction. In line with their general research on information robustness (Bergemann and Morris, 2005), Bergemann et al. (2016) point out that the designer is assumed to have perfect knowledge about type distribution and show the necessity of this assumption.

In this paper, we show that the full extraction auction proposed by Crémer and McLean (1985) is in general not robust towards outcome-based social preferences. In the standard case of an indivisible single unit good we conclude that purely selfish preferences are not only sufficient but also necessary for the existence of an ex post equilibrium where every bidder truthfully reveals her valuation. For the case of divisible goods or requiring only implementation in an Bayes Nash equilibrium we find that the existence depends on the intensity of the correlation as well as the intensity of the outcome-based social preferences.

We continue with the following structure. In the next Section 4.2 we introduce the environment and the utility function for our analysis. In Section 4.3 we characterize in detail the auction in our environment and state preliminary results. Afterwards we analyze in Section 4.4 the expected behavior of socially motivated bidders in the induced game by the auction and environment. In the last Section 4.5 we conclude.

4.2 Framework

4.2.1 The Environment

A single unit good is to be sold by an auction to one of two ex ante symmetric bidders $i \in \{1, 2\}$ with private valuation types $\theta_i \in \Theta_i = \{\theta^L, \theta^H\}$. We define the difference in valuation types as $\Delta\theta = \theta^H - \theta^L$. A valuation type profile is denoted by $\theta = (\theta_1, \theta_2) \in \Theta$. The ex ante distribution of valuation type profiles is given by $p : \Theta \to [0, 1]$.

We assume that valuation types between bidders are correlated. Interim, bidders update their belief about the other bidder's valuation type in the Bayesian sense. We denote the updated belief for each valuation type θ_i by $p(\cdot|\theta_i) : \Theta_j \to [0, 1]$. We also refer to them as conditional probabilities. The auctioneer is assumed to know the probability distribution over valuation types.

The set of feasible allocations A specifies which bidder receives the good and the according transfers paid to the auctioneer. A generic element $a \in A$ is a list (q_1, q_2, t_1, t_2). The allocation vector (q_1, q_2) states the amount of the good that is assigned to each bidder i. In general, it is restricted to $q_1 + q_2 \in [0, 1]$ and for the case of an indivisible good to $q_1 = 1 - q_2$ with $q_1 \in \{0, 1\}$. The associated transfers to the auctioneer for each bidder are $(t_1, t_2) \in \mathbb{R}^2$. An *environment* is described by $\mathbb{V} = (\{1, 2\}, \{\theta_1, \theta_2\}, p, A)$.

An *auction* $\Phi = (M, g)$ consists of a message profile $m = (m_1, m_2) \in M = M_1 \times M_2$ (bids) and an outcome function $g : M \to A$. We also write the list $g = (q_1^g, q_2^g, t_1^g, t_2^g)$. In the auction each bidder needs to report a bid $m_i \in M_i$ to the auction. Then the allocation function $q_i^g(m)$ represents whether the good is allocated to bidder i in the case of message profile m, and $t_i^g(m)$ the corresponding transfer for each bidder i to the seller.

For each bidder a strategy is the mapping $s_i : \Theta_i \to M_i$. The set of all such strategies of agent i is denoted $S_i = (s_i(\theta^L), s_i(\theta^H)$, and we write for the strategy profile $s = (s_1, s_2)$. For each valuation profile there exists a strategy vector $s(\theta) = (s_1(\theta_1), s_2(\theta_2))$. Given the auction Φ, we assume that bidders' payoffs are quasi-linear, i.e.

$$\pi_i((s_i(\theta_i), s_j(\theta_j)), \theta_i) = q_i(s_i(\theta_i), s_j(\theta_j))\ \theta_i + t_i(s_i(\theta_i), s_j(\theta_j)). \tag{4.1}$$

We denote an ex post payoff profile by

$$\pi(\theta_1, \theta_2) = (\pi_1((s_1(\theta_1), s_2(\theta_2)), \theta_1), \pi_2((s_2(\theta_2), s_1(\theta_1)), \theta_2)).$$

4.2.2 The Functional Form of the Utility

With the utility function we want to represent not only preferences over the own material payoff of bidders, but also acknowledge that bidders might compare their own material payoff with the other bidder's one. The utility function proposed by Charness and Rabin (2002) accounts for many kinds of outcome-based social preference identified in the experimental economic literature. With their model we are able to analyze the robustness of the *auction* by Crémer and McLean (1985) towards different kinds of outcome-based social preference.

In this paper, we follow Charness and Rabin (2002) functional form, but ignore reciprocal motives[1]. As Bartling et al. (2017) demonstrate in an experimental study that a bidder's social reference group includes other bidders but not the auctioneer, we allow only for outcome comparison among bidders. In detail, we define that

Definition 4.1. *The (ex post) utility function representing outcome-based social preferences is*

$$u_i(s_i, s_j, \theta, (\rho, \sigma)) = [\rho\ r + \sigma\ v]\ \pi_j((s_j, s_i), \theta_j) + \ [1 - \rho\ r - \sigma\ v]\ \pi_i((s_i, s_j), \theta_i) \tag{4.2}$$

where one distinguishes whether bidder i is payoff wise ahead ($\pi_i > \pi_j$) or behind ($\pi_i < \pi_j$). We follow the approach of the original parameterization and set

$$r = \begin{cases} 1 & \textit{if } \pi_i > \pi_j, \\ 0 & \textit{otherwise}, \end{cases}$$

$$v = \begin{cases} 1 & \textit{if } \pi_i < \pi_j, \\ 0 & \textit{otherwise}, \end{cases}$$

[1]See for instance Chapter 2 based on Bierbrauer and Netzer (2016) for a detailed discussion of the effect of intention-based preferences in mechanism design

such that $\rho \in [-1, 1]$ *illustrates the outcome-based social preferences when oneself is payoff wise ahead and* $\sigma \in [-1, 1]$ *otherwise.*

For the analysis we define *social preferences parameters* (ρ, σ) who represent empirical observations of outcome-based social preferences in the literature. Well known candidates are (purely) selfish ($\rho = \sigma = 0$), competitive ($\sigma \leq \rho < 0$), difference averse ($\sigma < 0 < \rho < 1$) and social-welfare oriented ($0 < \sigma \leq \rho \leq 1$) preferences. A detailed discussion of parameter combinations is provided by Charness and Rabin (2002).

In summary, a positive value of ρ or σ states that the bidder assigns a positive weight to the other bidder's payoff: the own utility increases in the payoff of the other bidder holding the own payoff fixed. The opposite is true for negative values of ρ or σ: the own utility decreases in the payoff of the other bidder. For the analysis of the Crémer-McLean auction we use the Charness and Rabin (2002) model to investigate under which outcome-based social preference types, represented by the social preference parameters (ρ, σ), a bidder prefers to deviate from the truth-telling strategy.

We also denote an ex post utility profile as $(u_i(s_i, s_j, \theta, (\rho, \sigma)), u_j(s_j, s_i, \theta, (\rho, \sigma)))$. Similar to an ex post payoff profile an ex post utility profile states the ex post utilities for both bidders given strategies s_i and s_j, the valuation type profile θ and social preference parameters (σ, ρ). If not otherwise stated, we refer to their ex post versions if we use the terms payoff or utility.

4.2.3 The Equilibrium Concepts

In order to conduct the analysis, we introduce *the ex post equilibrium* as our main equilibrium concept. For the analysis of interim behavior, we define the expected utility and the corresponding Bayes Nash equilibrium in our framework.

Intuitively, the ex post equilibrium requires that for each valuation type profile θ the corresponding strategies are ex post best responses to each other. Hence, for each realization of the valuation type profile, none of the two bidders would prefer to deviate to another strategy, if the private information of the other bidder is revealed. For example, truth-telling strategies form an ex post equilibrium if both bidders prefer to reveal their private informations for each valuation type of the other bidder.

In the case of private valuations and selfish preferences the truth-telling strategy profile is an ex post equilibrium of an incomplete information game induced by the direct mechanism if and only if it is an *equilibrium in dominant strategies* (see for instance Bergemann and Morris, 2005). The definition of the direct mechanism indicates the result: the type set and the message set are equal, i.e. each message represents one possible type. An equilibrium in dominant strategies requires that it remains a best response against each message of each type. But in an ex post equilibrium each bidder already prefers to reveal her information for any possible message of the other bidder.

This equivalence result holds not for all possible outcome-based social preferences as described in equation (4.2). If bidders have selfish preferences, only the own valuation matters for the utility and the consequential behavior. Hence, a bidder would be indifferent between a scenario where the true valuation type of the other bidder reporting a certain valuation is indeed low or high. But such a message might have different payoff consequences for these two valuation types of the other other bidder. Therefore, if a bidder now takes the other bidder's material well-being into account, these different payoff consequences matter for the utility and she is no longer indifferent between the two scenarios.

Nevertheless, the conditions for the existence of an ex post equilibrium remain a subset of the ones for the equilibrium in dominant strategies. Hence, the main theorem that shows that there does not exist an ex post equilibrium where each bidder reports truthful implies that there also does not exist an equilibrium in dominant strategies.

In Definition 4.2 we state this set of conditions for the ex post equilibrium. We define that

Definition 4.2. *An ex post equilibrium is a strategy profile $s^* = (s_1^*, s_2^*)$ such that for each valuation type profile $\theta \in \Theta$, for all $i, j \in \{1, 2\}$,*

$$u_i(s_i^*(\theta_i), s_j^*(\theta_j), \theta, (\rho, \sigma)) \geq u_i(m_i, s_j^*(\theta_j), \theta, (\rho, \sigma))$$

for all $m_i \in M_i$.

Given selfish preferences ex post equilibria are a subset of Bayes Nash equilibria. The later requires that bidders' strategies are optimal with respect to the expected utility formed by the belief over the other bidder's valuation types. We continue by defining the utility equivalence to expected payoffs: (interim) expected utility. That is the bidders expected utility from a strategy after she receives her own valuation type but is unaware about the realization for the other one. Formally,

Definition 4.3. *We define the interim expected utility representing outcome-based* *social preferences for bidder i's valuation type θ_i as*

$$\mathbb{E}_{\theta_j}\left[u_i(s_i(\theta_i), s_j(\theta_j), \theta, (\rho, \sigma))\right] = \sum_{\theta_j \in \Theta_j} p(\theta_j \mid \theta_i)\; u_i(s_i(\theta_i), s_j(\theta_j), \theta, (\rho, \sigma))$$

where $s_j(\theta_j)$ is the assigned strategy for valuation θ_j of bidder j.

Interim expected utility is the weighted sum of the ex post utilities. The weighting is with respect to the corresponding conditional probability of the valuation θ_j. The corresponding equilibrium concept, the Bayes Nash equilibrium, requires that for each bidder the strategy s_i^* is a best response to the other bidder's strategy s_j, weighted by the interim beliefs over potential valuations, i.e the conditional probability function $p(\cdot|\theta_i)$.

Formally, we define that

Definition 4.4. *A Bayes Nash equilibrium is a strategy profile $s^* = (s_1^*, s_2^*)$ such that for each bidder $i \in \{1, 2\}$, for all $\theta_i \in \Theta_i$,*

$$\mathbb{E}_{\theta_j}\left[u_i(s_i^*(\theta_i), s_j^*(\theta_j), \theta, (\rho, \sigma))\right] \geq \mathbb{E}_{\theta_j}\left[u_i(m_i, s_j^*(\theta_j), \theta, (\rho, \sigma))\right]$$

for all $m_i \in M_i$.

In contrast, in an ex post equilibrium the strategies were optimal with respect to the ex post payoff against any valuation type of the other bidder. Hence, in comparison to the ex post equilibrium the requirement is only in interim expectations, i.e. given the interim belief over valuation type of the other bidder. The ex post equilibria remain a subset of strategy profiles where the conditional probability of each valuation type is equal to one.

4.3 The Crémer-McLean Auction

We introduce now the Bayesian game induced by the auction proposed in Crémer and McLean (1985, 1988) to extract bidders' full expected surplus. We refer to such an auction as Φ^{CM}. Then we state preliminary results in Lemma 1 that imply consequences of the auction based on the correlation between valuation types. Φ^{CM} modifies the transfers of the SPA (Vickrey, 1961) by an additional lottery, which solely depends on the bid of the other bidder. We label this lottery in the following as *extraction lottery.*

The auction Φ^{CM}

Due to the revelation principle the message set is equal to the set of valuation types, i.e. $M_i = \{\theta^L, \theta^H\}$. The allocation function in Φ^{CM} is equal to the one in a SPA, i.e. for each $i \in \{1, 2\}$

$$q_i^{CM}(m_i, m_j) = \begin{cases} 1 & \text{if } m_i > m_j \\ 1 & \text{if } m_i = m_j \;\;\&\;\; l_i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

with $l_i \in \{0, 1\}$ the binary outcome of a fair lottery in the case of an indivisible good[2]. A fair lottery l assigns the same probability to all highest bidders to win the single unit good. In our case of two bidders, each of them has an equal probability of 0.5 to obtain the good. But, in comparison to the SPA, the transfer function for each bidder is modified such that for each $i \in \{1, 2\}$

$$t_i^{CM}(m_i, m_j) = -q_i(m_i, m_j)\; m_j + x(m_j) \tag{4.4}$$

where $-q_i(m_i, m_j)m_j$ is the transfer from the SPA and $x(m_j)$ is the proposed *extraction lottery* by Crémer and McLean (1985, 1988):

$$x(m_j) = \begin{cases} \frac{p(\theta^L|\theta^H)p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)p(\theta^H|\theta^H) - p(\theta^L|\theta^H)p(\theta^H|\theta^L)}\Delta\theta & \text{if } m_j = \theta^L, \\ -\frac{p(\theta^L|\theta^L)p(\theta^L|\theta^H)}{p(\theta^L|\theta^L)p(\theta^H|\theta^H) - p(\theta^L|\theta^H)p(\theta^H|\theta^L)}\Delta\theta & \text{if } m_j = \theta^H. \end{cases} \tag{4.5}$$

The extraction lottery

The *extraction lottery* depends only on the other bidders message m_j. Therefore the incentive compatibility in dominant strategies to report truthful from the SPA also holds for the Φ^{CM}. Bidder i can not directly affect the outcome of $x(m_j)$. The *extraction lottery* is designed such that in interim expectation each valuation type pays the amount that she would have received as an expected surplus in a SPA, i.e.

$$\mathbb{E}[\; x(m_j) \mid \theta^L] = 0, \tag{4.6}$$

$$\mathbb{E}[\; x(m_j) \mid \theta^H] = -p(\theta^L|\theta^H)\Delta\theta. \tag{4.7}$$

In Appendix 4.A we provide a detailed derivation of this *extraction lottery*. For the possibility to create such an *extraction lottery* the assumption of correlated valuation

[2] If the good is divisible the allocation function would be $q_i^{CM} = 0.5$ if $m_i = m_j$. We discuss this case in more detail in Proposition 4.1.

types is necessary. In that case the bidders' beliefs over the other bidder's valuation depends on whether the own valuation is low or high. This disagreement of valuation type distributions implements different interim valuations of the *extraction lottery* $x(m_j)$.

The payments $x(\theta^L)$ and $x(\theta^H)$ are constructed such that these interim valuations coincide with the expected surplus in a SPA. In combination with the expected payoff from the SPA, the *extraction lottery* yields therefore for both types an expected surplus of zero. We assume that conditions stated by Crémer and McLean (1985) are satisfied. To be concrete

Assumption 4.1. *We assume that the conditional probability matrix*

$$\begin{bmatrix} p(\theta^L|\theta^L) & p(\theta^H|\theta^L) \\ p(\theta^L|\theta^H) & p(\theta^H|\theta^H) \end{bmatrix} \tag{4.8}$$

has full rank.

Throughout the paper the considered correlation satisfies the condition in Assumption 4.1. We distinguish two cases of correlated valuation types. Either the two bidders are more or less similar in their valuation of the good. We refer to the first case as positive and to the later as negative correlation of valuation types. Formally,

Definition 4.5. *We define the following terms of type distributions:*
(a) A type distribution is positive correlated if $p(\theta^L|\theta^L)p(\theta^H|\theta^H) > p(\theta^H|\theta^L)p(\theta^L|\theta^H)$.
(b) A type distribution is negative correlated if $p(\theta^L|\theta^L)p(\theta^H|\theta^H) < p(\theta^H|\theta^L)p(\theta^L|\theta^H)$.
(c) A type distribution is perfect correlated if either $p(\theta^L|\theta^L) = p(\theta^H|\theta^H) = 1$ *or* $p(\theta^L|\theta^H) = p(\theta^H|\theta^L) = 1$.

Lemma 1. *We state for each correlation the following consequences for the extraction lottery:*
(a) Suppose a positive but not perfect correlated type distribution. Then $x(\theta^L) > 0 > x(\theta^H)$.
(b) Suppose a negative but not perfect correlated type distribution. Then $x(\theta^L) < -\Delta\theta$ *and* $x(\theta^H) > 0$.

Proof. Suppose the type distribution is not perfect correlated. Then by Definition 4.5 (c) $p(\theta^L|\theta^H)p(\theta^H|\theta^L) > 0$ and $p(\theta^L|\theta^L)p(\theta^L|\theta^H) > 0$.
(a) Suppose the type distribution is positive correlated. Then by Definition 4.5 (a) $p(\theta^L|\theta^L)p(\theta^H|\theta^H) > p(\theta^H|\theta^L)p(\theta^L|\theta^H)$ which implies that $p(\theta^L|\theta^L)p(\theta^H|\theta^H) - p(\theta^H|\theta^L)p(\theta^L|\theta^H) > 0$. For not perfect correlation the payments calculated in (4.5)

are $x(\theta^L) > 0$ and $x(\theta^H) < 0$.
(b) Suppose the type distribution is negative correlated. Then by Definition 4.5 (b) $p(\theta^L|\theta^L)p(\theta^H|\theta^H) < p(\theta^H|\theta^L)p(\theta^L|\theta^H)$ which implies that $p(\theta^L|\theta^L)p(\theta^H|\theta^H) - p(\theta^H|\theta^L)p(\theta^L|\theta^H) < 0$. For not perfect correlation the payments calculated in (4.5) are $x(\theta^L) < -\Delta\theta$ and $x(\theta^H) > 0$. □

In Figure 4.1 we display how the payments in the *extraction lottery* depend on the correlation. We illustrate this correlation by $p(\theta^L|\theta^L)p(\theta^H|\theta^H) - p(\theta^H|\theta^L)p(\theta^L|\theta^H) \in [-1, 1]$.

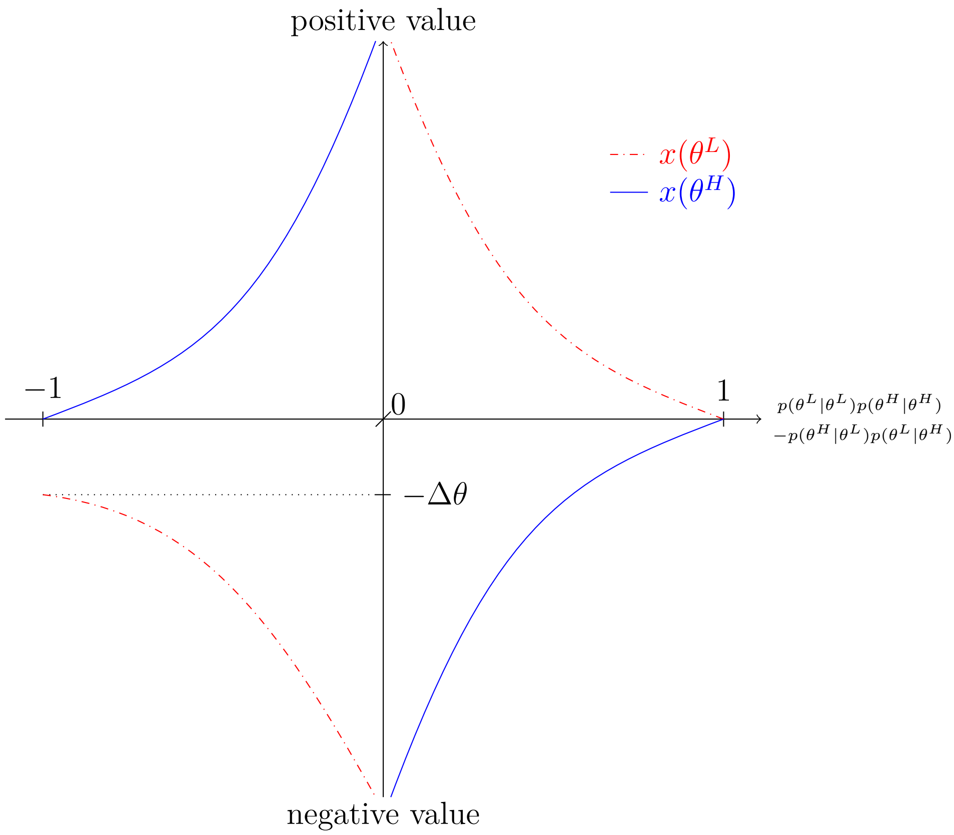


Figure 4.1: Payment dependence on type reports for two types

Figure 4.1 illustrates that a stronger (positive or negative) correlation leads to a lower difference in absolute terms of payments. If the correlation is stronger, then the disagreement of valuation types in their beliefs about the other bidder's type is more pronounced. Therefore, lower monetary incentives are sufficient to satisfy equation (4.6) and (4.7).

The set of conditions for incentive compatibility

The main focus in the next section is the interval of (ρ, σ) in $[-1, 1]^2$ such that the truth-telling strategy profile s^{TT} forms an ex post equilibrium in the Bayesian game induced by the auction Φ^{CM} in the environment $\mathbb{V}$. There exist four possible ex post valuation type profiles that need to be considered: (θ^L, θ^L), (θ^L, θ^H), (θ^H, θ^L), and (θ^H, θ^H).

If both bidders report the same message, i.e. $m_1 = m_2$, the single unit good is allocated by the fair lottery l. Remember that in the fair lottery both bidders have an equal probability to obtain the good. The good is assumed to be indivisible. That means if one bidder receives the good, the other one does not. Note that the allocation $q_i \in \{0, 1\}$ has direct effect on the transfer function t_i^{CM}. Hence, given $m_1 = m_2$, there exist two ex post payoff profiles in each of the ex post valuation type profiles.

Table 4.1 illustrates the ex post payoff profiles π for each valuation type profile θ. The top row states the payoff profile if the row bidder wins the fair lottery, the bottom one when she loses. If both payoff profiles are equivalent, only one row is represented. The ex post payoff profiles given that both bidders report truthful are underlined.

(θ^L, θ^L)	θ^L	θ^H
θ^L	$\underline{x(\theta^L), x(\theta^L)}$	$x(\theta^H), x(\theta^L)$
θ^H	$x(\theta^L), x(\theta^H)$	$-\Delta\theta + x(\theta^H), x(\theta^H)$ $x(\theta^H), -\Delta\theta + x(\theta^H)$

(θ^L, θ^H)	θ^L	θ^H
θ^L	$x(\theta^L), x(\theta^L)$ $x(\theta^L), \Delta\theta + x(\theta^L)$	$\underline{x(\theta^H), \Delta\theta + x(\theta^L)}$
θ^H	$x(\theta^L), x(\theta^H)$	$-\Delta\theta + x(\theta^H), x(\theta^H)$ $x(\theta^H), x(\theta^H)$

(θ^H, θ^L)	θ^L	θ^H
θ^L	$\Delta\theta + x(\theta^L), x(\theta^L)$ $x(\theta^L), x(\theta^L)$	$x(\theta^H), x(\theta^L)$
θ^H	$\underline{\Delta\theta + x(\theta^L), x(\theta^H)}$	$x(\theta^H), x(\theta^H)$ $x(\theta^H), -\Delta\theta + x(\theta^H)$

(θ^H, θ^H)	θ^L	θ^H
θ^L	$\Delta\theta + x(\theta^L), x(\theta^L)$ $x(\theta^L), \Delta\theta + x(\theta^L)$	$x(\theta^H), \Delta\theta + x(\theta^L)$
θ^H	$\Delta\theta + x(\theta^L), x(\theta^H)$	$\underline{x(\theta^H), x(\theta^H)}$

Table 4.1: Payoff tables depending on the valuation type profile θ

4.4 Behavioral Predictions

In this section, we investigate the behavioral predictions of social preferences in the Bayesian game induced by the full extraction auction Φ^{CM}. We test the robustness by asking for which social preferences parameters (ρ, σ) in the interval $[-1, 1]^2$ the truth-telling strategy profile s^{TT} forms an equilibrium.

We start with the most restrictive equilibrium concept of an ex post equilibrium for indivisible goods. We relax then the conditions such that we ask for the interval that forms an ex post equilibrium for divisible goods and in the last part a Bayes Nash equilibrium. We conclude that the decrease of the conditions increases the interval in which the truth-telling strategy profile is an equilibrium but show that all conditions are never satisfied for the full set of social preferences parameters $[-1, 1] \times [-1, 1]$.

Ex post equilibrium (indivisible good)

To ensure that the truth-telling strategy profile forms an ex post equilibrium eight incentive compatibility constraints need to be satisfied. Remember that in contrast to the original analysis we allow bidders behavior to be motivated by outcome-based social preferences. In Appendix 4.B we provide a detailed analysis on the conditions for the social preference parameters (ρ, σ) to ensure that the truth-telling strategy profile s^{TT} indeed forms an ex post equilibrium. We find that

Theorem 4.1. *The truth-telling strategy profile s^{TT} forms an ex-post equilibrium in the Bayesian game induced by Φ^{CM} for the indivisible good case if and only if the bidders have (pure) selfish preferences, i.e. $\rho = \sigma = 0$.*

Proof. See conditions in Appendix 4.B. □

The indivisibility of the single unit good leads to an allocation choice by the fair lottery with two potential ex post allocations: either bidder i receives or does not receive the indivisible good. The intuition behind Theorem 4.1 is that in at least one of these two possibilities the incentive compatibility constraint is binding for each valuation type in all valuation type profiles. Also, in Table 4.1 we see that for each complete information game a deviation from the truth-telling strategy equilibrium yields for every valuation type in one of the two possibilities (winning or losing the fair lottery) the same ex post payoff.

Why is that the case? Remember that the transfer due to the extraction lottery $x(s_j)$ is independent from the allocation function $q_i(s)$. If both bidders report the same strategy, then the allocation is randomly determined by the fair lottery. This means that the allocation in the deviation must be equal to one of the two allocations by the fair lottery. The additional *extraction lottery* in Φ^{CM} is such that one outcome needs to be positive while the other one is negative. This means that the deviation of one valuation type decreases the payoff of the other bidder while the one for the other increases it.

In Figure 4.2 we illustrate the consequences of these deviations for the payoff profiles. The payoff profile for each valuation type profile where both bidders report truthful are underlined (and red). The payoff profiles if bidder i deviates are illustrated with their according valuation type which is not underlined. The arrow points from the equilibrium payoff profile to the deviating payoff profile. For instance, in the case of positive correlation, there are two occasions where bidder i receives a positive ex post payoff ($\pi_i(\theta_i, \theta_j) > 0$) and two where she receives a negative one ($\pi_i(\theta_i, \theta_j) < 0$). In the case bidder i has a high valuation while the other bidder has a low valuation, i.e. (θ^H, θ^L), bidder j receives a negative payoff. The arrow shows that a deviation by bidder i increases the ex post payoff of bidder j from x^H to x^L without affecting the own payoff which is in both cases $\Delta\theta + x^L$. Hence, with a deviation bidder i has a costless opportunity to decrease the inequality between the two bidders.

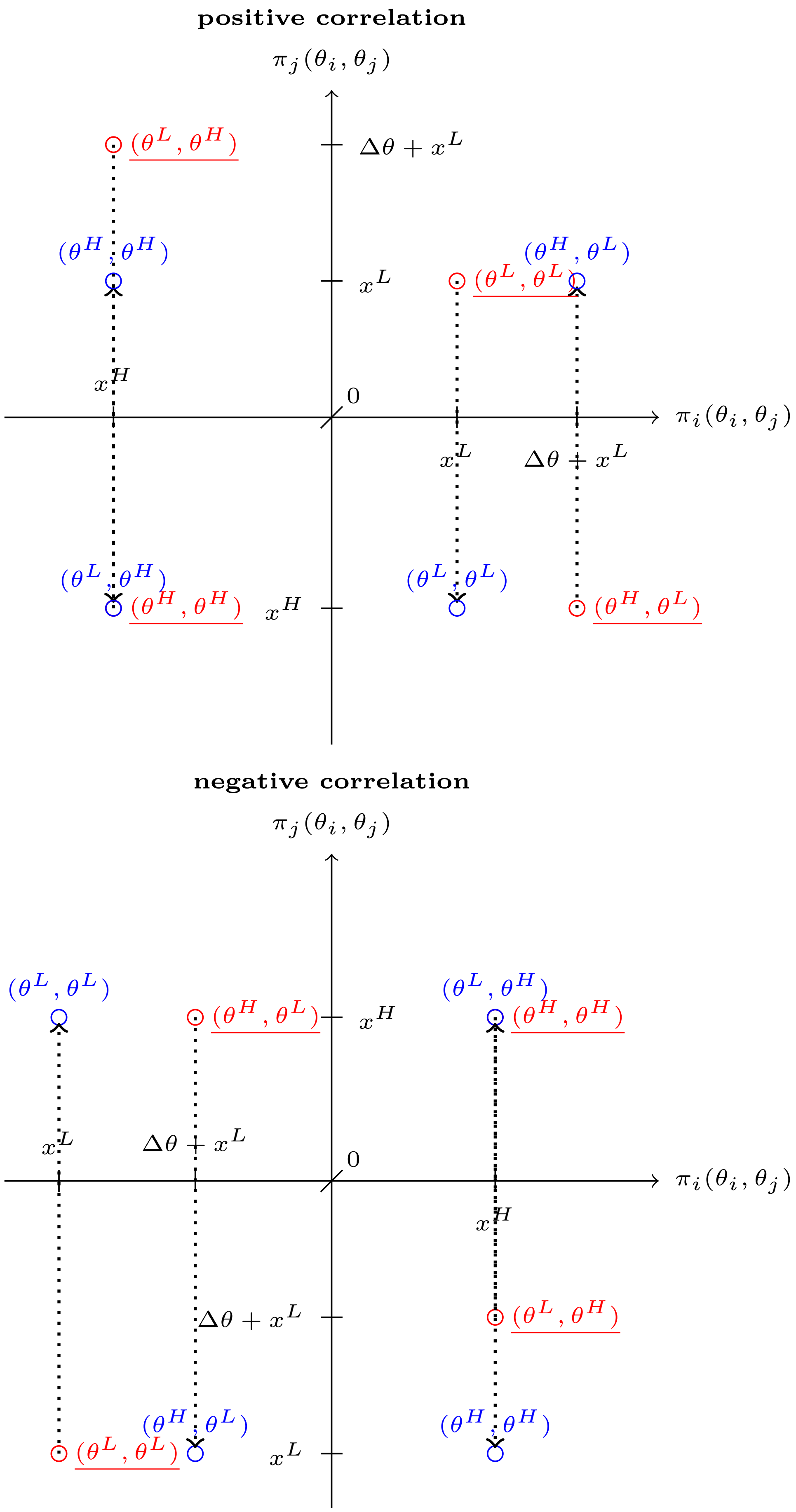


Figure 4.2: Ex post payoff profiles for truthful revelation

Note first that if both bidders truthfully reveal their valuation types, no valuation type receives an ex post payoff of zero in each of the both types of correlation. Hence, the auction never extracts the full ex post surplus of at least one of the two bidders.

Given that both bidders have the same valuation type, their ex post payoffs are equal. A unilateral deviation either increases or decreases the other bidder's payoff. The concrete consequences depend on the correlation and the valuation types. But for both types of correlation, once it allows the deviant to decrease the other bidder's payoff and once to increase it, without affecting the own payoff. In order to ensure that such a deviation is not preferred, the outcome-based social preferences should not include that one prefers that the other receives more payoff when one is behind ($\sigma > 0$) and at the same time less payoff when one is ahead ($\rho < 0$).

The intuition is similar for the case of different valuation types. Here, the ex post payoffs are unequal if both bidders truthfully reveal. Indeed, due to the extraction lottery there exists always one positive and one negative payoff. A unilateral deviation enables a bidder to reduce this asymmetry in the payoff dimension. If she is behind she can reduce the other bidder's payoff. In order to ensure not such a deviation we need to reduce the outcome-based social preferences such that she prefers not to decrease the inequality if she is behind ($\sigma < 0$). If she is ahead she can increase the other bidder's payoff. Therefore we need to restrict the parameters such that she does not prefer to decrease the inequality if she is ahead ($\rho > 0$). Generally, if valuations are unequal, difference averse bidders prefer to deviate.

The possibility to affect costless the other bidder's payoff implies (even in the case of weak pronounced outcome-based social preferences) a deviation from the truthful report.

Ex post equilibrium (divisible good)

The case when bidders have different valuation types depends crucially on our definition of indivisible goods. The necessary tie breaking rule creates additional ex post incentive-constraint conditions which need to be satisfied. More accurate, there are two outcomes of the allocation function, i.e. $q_i \in \{0, 1\}$, if $m_1 = m_2$. That the own payoff is not affected by a deviation is only true for one of the two outcomes. Hence, it cannot be true for $q_i = 0.5$ because the deviating payoff is then equal to the average of the two payoffs considered for Theorem 1. Indeed, we find that

Proposition 4.1. *For the case of a divisible good, i.e.* $q_i(m) = 0.5$ *if* $m_1 = m_2$*, the truth-telling strategy profile* s^{TT} *is an ex post equilibrium in the Bayesian game induced by* Φ^{CM} *for a certain range of outcome-based social preferences parameters. These parameters coincide with (weak) differences aversion and are given for a*

- *positive correlation by*

$\rho \in \left[0, \frac{0.5\Delta\theta}{[0.5\Delta\theta + x(\theta^L) - x(\theta^H)}\right]$,

$\sigma \in \left[-\frac{0.5\Delta\theta}{[0.5\Delta\theta + x(\theta^L) - x(\theta^H)}, 0\right]$.

- *negative correlation by*

$\rho \in \left[0, \frac{(1-\sigma)0.5\Delta\theta}{x(\theta^H) - x(\theta^L) - \Delta\theta}\right]$,

$\sigma \in \left[-\frac{(1+\rho)0.5\Delta\theta}{x(\theta^H) - x(\theta^L) - \Delta\theta}, 0\right]$.

Proof. See calculations in Appendix 4.B and set $q(\theta^L, \theta^L) = q(\theta^H, \theta^H) = 0.5$. □

Intuitively, the proposition sates that for these parameters the bidders' best response to the truth-telling strategy of the other bidder remains the truth-telling strategy. The intuition for this possibility is due to the effect that the allocation function cannot remain constant in a deviation from the truthful report.

In the case of asymmetric valuations it is no longer possible that the own payoff can be the same for two different messages. That means there is no costless possibility to affect the other bidder's payoff. In order to affect the payoff of the other bidder, the bidder would need to change the own payoff. In other words, the bidder faces a trade-off. If the outcome-based social preferences are not sufficiently pronounced, she remains from that possibility and prefers to reveal her valuation type.

This is not true for the case of symmetric valuation where it is still possible to affect the other bidder's payoff without any cost. The reason is that the costless deviation in terms of the own payoff is possible for a bidder for both payoff profiles. Hence, it must be true for any convex combination of these two payoff profiles, including $q_i(m) = 0.5$.

Note that we have seen in Figure 4.1 that the stronger the correlation between the two bidders, the smaller the absolute difference between lottery payments $|x(\theta^H) - x(\theta^L)|$. Hence, we find that the intensity of correlation directly affects whether the truth-telling equilibrium exists if the goods is divisible and bidders are difference averse. Nevertheless, other social preference types such as competitive or social-welfare oriented preferences would still deviate from the truth-telling strategy ex post for any intensity of strict correlation between valuation types.

Bayes Nash equilibrium

In the next step we reduce the amount of incentive constraints even further and investigate interim prediction of bidders behavior. In other words, we ask for which range of social preference parameters the truth-telling strategy profile forms a Bayes Nash equilibrium. In general, we find that

Proposition 4.2. *Whether the truth-telling strategy profile s^{TT} is a Bayes Nash equilibrium depends on the correlation between valuation types and outcome-based social preferences parameters. In detail, the interval for σ such that there exists a Bayes Nash equilibrium is given for*

- *positive correlation by*

$$\sigma \in \left[-\frac{p(\theta^L|\theta^L)}{p(\theta^H|\theta^L)}\frac{x^L - x^H}{x^L - x^H + 0.5\Delta\theta}\rho - \frac{0.5\Delta\theta}{x^L - x^H + 0.5\Delta\theta}, -\frac{p(\theta^L|\theta^L)}{p(\theta^H|\theta^L)}\frac{x^L - x^H + 0.5\Delta\theta}{x^L - x^H + \Delta\theta}\rho + \frac{p(\theta^L|\theta^L)}{p(\theta^H|\theta^L)}\frac{0.5\Delta\theta}{x^L - x^H + \Delta\theta}\right]$$

- *negative correlation by*

$$\sigma \in \left[-\frac{p(\theta^L|\theta^H)0.5\Delta\theta + p(\theta^H|\theta^H)[x^H - x^L - \Delta\theta]}{p(\theta^L|\theta^H)[x^H - x^L + \Delta\theta]}\rho - \frac{0.5\Delta\theta}{x^H - x^L + \Delta\theta}, \frac{p(\theta^H|\theta^L)[\rho(x^L - x^H + \Delta\theta) + 0.5\Delta\theta]}{p(\theta^L|\theta^L)[x^H - x^L] + p(\theta^H|\theta^L)0.5\Delta\theta}\right].$$

In general, the truth-telling strategy profile forms not a Bayes Nash equilibrium for all $(\sigma, \rho) \in [-1, 1]^2$.

Proof. By counterexample.
(1) Suppose positive correlation and $\rho = 0$. Then the lower bound of the interval for σ such that the truth-telling strategy forms a Bayes Nash equilibrium is $-\frac{0.5\Delta\theta}{x^L - x^H + 0.5\Delta\theta}$. This is strictly greater than -1 for any but perfect positive correlation.
(2) Suppose negative correlation and $\rho = 0$. Then, the lower bound of the interval for σ such that the truth-telling strategy forms a Bayes Nash equilibrium is $-\frac{0.5\Delta\theta}{x^H - x^L + \Delta\theta}$. This is strictly greater than -1 for any negative correlation. □

In comparison to the ex post equilibrium in the case of divisible goods, interim there are no deviations such that the own payoff remains the same. Hence in the case of non-selfish preferences there always exist a trade-off between affecting the own and the other bidder's payoff.

Remember that the intensity of the correlation affects the absolute difference between payoffs in the extraction lottery. Hence, whether the truth-telling strategy profile forms a Bayes Nash equilibrium also depends on the interaction between two factors. First, on the intensity of the outcome-based social preferences and, second, on the intensity of the correlation of valuation types. This interaction is non-trivial and the general insight we provide here is that the auction Φ^{CM} is not robust against outcome-based social preference in the sense that the truth-telling strategy profile does not form a Bayes Nash equilibrium for all potential outcome-based social preference parameter combinations (σ, ρ).

4.5 Conclusion

In this chapter we investigate whether the first best implementation in environments with correlated type distribution is robust towards outcome-based social preferences. We find that this not the case in general. If the single unit good is indivisible, we even conclude that selfish preferences are not only a sufficient but also a necessary condition for the existence of the truth-telling ex post equilibrium.

Hence the auctioneer's expected profit depends on the distribution of social preferences among bidders. For certain distributions she might be even better off than under symmetric information. Whether the auctioneer would like to use the extraction lottery depends on her belief about the distribution of outcome-based social preferences.

It is worth to notice that this conclusion neglects interim voluntarily participation. Most of the behavioral critics so far concentrates on that point. For instance, the voluntary participation constraint of the auction by Crémer and McLean (1985) is not satisfied under risk aversion or limited liability. In our setting interim participation depends on the bidders' social preferences and their beliefs about the distribution of social preferences in the population.

In the next chapter we specify the optimal auction under an externality freeness property. This property allows for social robust implementation and we report a comparison with optimal auction under independent valuation type distribution. We believe that this is a first step to deal with the non-robustness property of first best implementation, if it is in the interest of the auctioneer.

4.A Derivation of the Extraction Lottery

$$\mathbb{E}[x(m_j) \mid \theta^L] = 0 \tag{4.9}$$
$$\mathbb{E}[x(m_j) \mid \theta^H] = -p(\theta^L|\theta^H)\Delta\theta \tag{4.10}$$

Rewrite (4.9) :

$$\begin{aligned}
(4.9) &\Leftrightarrow p(\theta^L|\theta^L)\ x(\theta^L) + p(\theta^H|\theta^L)\ x(\theta^H) = 0 \\
&\Leftrightarrow x(\theta^L) = -\frac{p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)} x(\theta^H)
\end{aligned}$$

Plug into (4.10) :

$$\begin{aligned}
(4.10) &\Leftrightarrow p(\theta^L|\theta^H)\ x(\theta^L) + p(\theta^H|\theta^H)\ x(\theta^H) = -p(\theta^L|\theta^H)\Delta\theta \\
&\Leftrightarrow p(\theta^L|\theta^H)\left[-\frac{p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)} x(\theta^H)\right] + p(\theta^H|\theta^H)\ x(\theta^H) = -p(\theta^L|\theta^H)\Delta\theta \\
&\Leftrightarrow\ - p(\theta^L|\theta^H)\ p(\theta^H|\theta^L) x(\theta^H) + p(\theta^L|\theta^L)\ p(\theta^H|\theta^H)\ x(\theta^H) = -p(\theta^L|\theta^L)\ p(\theta^L|\theta^H)\Delta\theta \\
&\Leftrightarrow [p(\theta^L|\theta^L)\ p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\ p(\theta^H|\theta^L)]\ x(\theta^H) = -p(\theta^L|\theta^L)\ p(\theta^L|\theta^H)\Delta\theta \\
&\Leftrightarrow x(\theta^H) = -\frac{p(\theta^L|\theta^L)\ p(\theta^L|\theta^H)}{p(\theta^L|\theta^L)\ p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\ p(\theta^H|\theta^L)}\Delta\theta
\end{aligned}$$

Plug into $x(\theta^L)$:

$$\begin{aligned}
x(\theta^L) &= -\frac{p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)} x(\theta^H) \\
\Leftrightarrow x(\theta^L) &= -\frac{p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)}\left[-\frac{p(\theta^L|\theta^L)\ p(\theta^L|\theta^H)}{p(\theta^L|\theta^L)\ p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\ p(\theta^H|\theta^L)}\Delta\theta\right] \\
\Leftrightarrow x(\theta^L) &= \frac{p(\theta^L|\theta^H)\ p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)\ p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\ p(\theta^H|\theta^L)}\Delta\theta
\end{aligned}$$

4.B Ex Post Incentive Constraints

Suppose first strict **positive correlation** $[p(\theta^L|\theta^L)p(\theta^H|\theta^H) > p(\theta^L|\theta^H)p(\theta^H|\theta^L)]$ such that $x(\theta^L) > 0$ and $x(\theta^H) < 0$:

Assume $\theta = (\theta^L, \theta^L)$:

$$\begin{aligned} \theta^L \succeq \theta^H &\Leftrightarrow x(\theta^L) \geq \rho\, x(\theta^H) + (1-\rho)\, x(\theta^L) \\ &\Leftrightarrow 0 \geq \rho\, [x(\theta^H) - x(\theta^L)] \\ &\Leftrightarrow \rho \geq 0 \end{aligned} \tag{4.11}$$

Assume $\theta = (\theta^H, \theta^H)$:

$$\begin{aligned} \theta^H \succeq \theta^L &\Leftrightarrow x(\theta^H) \geq \sigma\, [\Delta\theta + x(\theta^L)] + (1-\sigma)\, x(\theta^H) \\ &\Leftrightarrow 0 \geq \sigma\, [\Delta\theta + x(\theta^L) - x(\theta^H)] \\ &\Leftrightarrow \sigma \leq 0 \end{aligned} \tag{4.12}$$

Assume $\theta = (\theta^L, \theta^H)$:

For θ_L:

$$\begin{aligned} \theta^L \succeq \theta^H &\Leftrightarrow \sigma[\Delta\theta + x(\theta^L)] + (1-\sigma)x(\theta^H) \geq \sigma([1 - q_i(\theta^H, \theta^H)]0 + x(\theta^H)) + (1-\sigma)[q_i(\theta^H, \theta^H)(-\Delta\theta) \\ &\Leftrightarrow \sigma \geq -\frac{q_i(\theta^H, \theta^H)\Delta\theta}{[1 - q_i(\theta^H, \theta^H)]\Delta\theta + x(\theta^L) - x(\theta^H)} \end{aligned} \tag{4.13}$$

In the case of indivisible goods $q \in \{0, 1\}$:

$$q_i(\theta^H, \theta^H) = 0 \Leftrightarrow \sigma \geq 0 \tag{4.14}$$

$$q_i(\theta^H, \theta^H) = 1 \Leftrightarrow \sigma \geq -\frac{\Delta\theta}{x(\theta^L) - x(\theta^H)} \tag{4.15}$$

For θ_H:

$$\begin{aligned} \theta^H \succeq \theta^L &\Leftrightarrow \rho x(\theta^H) + (1-\rho)[\Delta\theta + x(\theta^L)] \geq \rho x(\theta^L) + (1-\rho)[x(\theta^L) + q_i(\theta^L, \theta^L)\Delta\theta] \\ &\Leftrightarrow \rho \leq \frac{(1 - q_i(\theta^L, \theta^L))\Delta\theta}{(1 - q_i(\theta^L, \theta^L))\Delta\theta + x(\theta^L) - x(\theta^H)} \end{aligned} \tag{4.16}$$

In the case of indivisible goods $q \in \{0, 1\}$:

$$q_i(\theta^L, \theta^L) = 0 \Leftrightarrow \rho \leq \frac{\Delta\theta}{\Delta\theta + x(\theta^L) - x(\theta^H)} \tag{4.17}$$

$$q_i(\theta^L, \theta^L) = 1 \Leftrightarrow \rho \leq 0 \tag{4.18}$$

We conclude the only parameter combination satisfying all conditions is $\rho = \sigma = 0$.

Suppose now strict **negative correlation** $[p(\theta^L|\theta^L)p(\theta^H|\theta^H) < p(\theta^L|\theta^H)p(\theta^H|\theta^L)]$ such that $x(\theta^L)(\theta^L) < -\Delta\theta$ and $x(\theta^L)(\theta^H) > 0$:

Assume $\theta = (\theta^L, \theta^L)$:

$$\begin{aligned} \theta^L \succeq \theta^H &\Leftrightarrow x(\theta^L) \geq \sigma\ x(\theta^H) + (1-\sigma)\ x(\theta^L) \\ &\Leftrightarrow 0 \geq \sigma\ [x(\theta^H) - x(\theta^L)] \\ &\Leftrightarrow \sigma \leq 0 \end{aligned} \tag{4.19}$$

Assume $\theta = (\theta^H, \theta^H)$:

$$\begin{aligned} \theta^H \succeq \theta^L &\Leftrightarrow x(\theta^H) \geq \rho\ [\Delta\theta + x(\theta^L)] + (1-\rho)\ x(\theta^H) \\ &\Leftrightarrow 0 \geq \rho\ [\Delta\theta + x(\theta^L) - x(\theta^H)] \\ &\Leftrightarrow \rho \geq 0 \end{aligned} \tag{4.20}$$

Assume $\theta = (\theta^L, \theta^H)$:

For θ_L:

$$\begin{aligned} \theta^L \succeq \theta^H &\Leftrightarrow \sigma\rho[\Delta\theta + x(\theta^L)] + (1-\rho)x(\theta^H) \geq \sigma x(\theta^H) + (1-\sigma)[-q_i(\theta^H,\theta^H)\Delta\theta + x(\theta^H)] \\ &\Leftrightarrow \rho \leq \frac{(1-\sigma)q_i(\theta^H,\theta^H)\Delta\theta}{x(\theta^H) - \Delta\theta - x(\theta^L)} \end{aligned} \tag{4.21}$$

In the case of indivisible goods $q \in \{0, 1\}$:

$$q_i(\theta^L,\theta^L) = 0 \Leftrightarrow \rho \leq 0 \tag{4.22}$$

$$q_i(\theta^L,\theta^L) = 1 \Leftrightarrow \rho \leq \frac{(1-\sigma)\Delta\theta}{x(\theta^H) - \Delta\theta - x(\theta^L)} \tag{4.23}$$

For θ_H:

$$\begin{aligned} \theta^H \succeq \theta^L &\Leftrightarrow \sigma x(\theta^H) + (1-\sigma)[\Delta\theta + x(\theta^L)] \geq \rho x(\theta^L) + (1-\rho)[x(\theta^L) + q_i(\theta^L,\theta^L)\Delta\theta] \\ &\Leftrightarrow \sigma \geq \frac{(1-\rho)q_i(\theta^L,\theta^L)\Delta\theta - \Delta\theta}{x(\theta^H) - x(\theta^L) - \Delta\theta} \end{aligned} \tag{4.24}$$

In the case of indivisible goods $q \in \{0,1\}$:

$$q_i(\theta^L, \theta^L) = 0 \Leftrightarrow \sigma \geq -\frac{\Delta\theta}{x(\theta^H) - x(\theta^L) - \Delta\theta} \tag{4.25}$$

$$q_i(\theta^L, \theta^L) = 1 \Leftrightarrow \sigma \geq -\rho\frac{\Delta\theta}{x(\theta^H) - x(\theta^L) - \Delta\theta} \tag{4.26}$$

From equation (4.20) and (4.22) we can conclude that $\rho = 0$. Plug this into equation (4.26) and we receive

$$\sigma \geq 0 \tag{4.27}$$

such that we can conclude in combination with equation (4.19) that $\sigma = 0$.

4.C Ex Post Utility Function

(θ^L, θ^L)	θ^L	θ^H
θ^L	x^L	$\sigma x^L + (1-\sigma)x^H$
θ^H	$\rho x^H + (1-\rho)x^L$	$\sigma x^H + (1-\sigma)[x^H - \Delta\theta]$ $\rho[x^H - \Delta\theta] + (1-\rho)x^H$

(θ^L, θ^H)	θ^L	θ^H
θ^L	x^L $\sigma[x^L + \Delta\theta] + (1-\sigma)x^L$	$\sigma[x^L + \Delta\theta] + (1-\sigma)x^H$
θ^H	$\rho x^H + (1-\rho)x^L$	$\sigma x^H + (1-\sigma)[x^H - \Delta\theta]$ x^H

(θ^H, θ^L)	θ^L	θ^H
θ^L	$\rho x^L + (1-\rho)[x^L + \Delta\theta]$ x^L	$\sigma x^L + (1-\sigma)x^H$
θ^H	$\rho x^H + (1-\rho)[x^L + \Delta\theta]$	x^H $\rho[x^H - \Delta\theta] + (1-\rho)x^H$

(θ^H, θ^H)	θ^L	θ^H
θ^L	$\rho x^L + (1-\rho)[x^L + \Delta\theta]$ $\sigma[x^L + \Delta\theta] + (1-\sigma)x^L$	$\sigma[x^L + \Delta\theta] + (1-\sigma)x^H$
θ^H	$\rho x^H + (1-\rho)[x^L + \Delta\theta]$	x^H

Table 4.2: Ex post utility function for positive correlation

(θ^L, θ^L)	θ^L	θ^H
θ^L	x^L	$\rho x^L + (1-\rho)x^H$
θ^H	$\sigma x^H + (1-\sigma)x^L$	$\sigma x^H + (1-\sigma)[x^H - \Delta\theta]$ $\rho[x^H - \Delta\theta] + (1-\rho)x^H$

(θ^L, θ^H)	θ^L	θ^H
θ^L	x^L $\sigma[x^L + \Delta\theta] + (1-\sigma)x^L$	$\rho[x^L + \Delta\theta] + (1-\rho)x^H$
θ^H	$\sigma x^H + (1-\sigma)x^L$	$\sigma x^H + (1-\sigma)[x^H - \Delta\theta]$ x^H

(θ^H, θ^L)	θ^L	θ^H
θ^L	$\rho x^L + (1-\rho)[x^L + \Delta\theta]$ x^L	$\rho x^L + (1-\rho)x^H$
θ^H	$\sigma x^H + (1-\sigma)[x^L + \Delta\theta]$	x^H $\rho[x^H - \Delta\theta] + (1-\rho)x^H$

(θ^H, θ^H)	θ^L	θ^H
θ^L	$\rho x^L + (1-\rho)[x^L + \Delta\theta]$ $\sigma[x^L + \Delta\theta] + (1-\sigma)x^L$	$\rho[x^L + \Delta\theta] + (1-\rho)x^H$
θ^H	$\sigma x^H + (1-\sigma)[x^L + \Delta\theta]$	x^H

Table 4.3: Ex post utility function for negative correlation

Chapter 5

Social Robust Auctions: The Case of Correlated Valuations

5.1 Introduction

The economic literature observes that social preferences can affect the behavior of economic agents within strategic interaction. An early illustration is the ultimatum game, which models bargaining situations. There social preferences explain at least partially for experimental deviations from the unique subgame perfect equilibrium (Güth et al., 1982). In the last decades many theoretical models incorporate outcome- and intention-based social preferences to improve the predictability in many strategic environments.

The environment we consider here is the allocation of a single unit good with an auction. In a laboratory experiment Andreoni et al. (2007) find deviations from the dominant strategy equilibrium in a second price auction and attribute it partially to spiteful preferences. Bartling and Netzer (2016) show that if other bidders are simulated by a computer program, behavior in their auction design is more consistent with selfish preference predictions.

One of the major challenges for the design of optimal auctions is that the intensity of social preferences is heterogeneous within the population and it depends on specific elements like framing. One solution to this heterogeneity problem is provided by Bierbrauer and Netzer (2016) who introduce an externality-freeness condition which ensures social robustness of social choice functions. The authors show that for uncorrelated private information types in expectation the same profit is achievable by a modified social choice function.

In this chapter we study how this externality-freeness condition affects the optimal auction when private informations over valuation types are correlated. Under selfish preferences the auctioneer has the possibility to achieve the same expected profit as under public information (Crémer and McLean, 1985, 1988). In the previous chapter we demonstrate the dependence of this result on the assumption of selfish preferences. This result suggests social preferences might have wide consequences in the environment. Therefore, we determine how the externality-freeness condition affects the set of possible allocations.

We derive the optimal auction under Bayesian and ex post incentive compatibility. The characterization enables for two different comparison. First, on the effect of introducing social preferences in the realm of correlated valuations, i.e. we compare the predicted outcome of the auctions derived here with the one by Crémer and McLean (1985, 1988). We find that if externality-freeness is required, it is no longer possible to implement the same expected payoff as under public information. Hence, the additional condition reduces the set of implementable social choice functions in correlated environments. In a second step, we compare the relation of the optimal auction under correlated valuation with the auction that requires the externality-freeness condition under the independence assumption (Bartling and Netzer, 2016).

In the case of Bayes Nash incentive compatibility, we compare our result to the externality-robust auction by Bartling and Netzer (2016). We find a similar transfer function, but due to the correlated valuation assumption it depends on the conditional probability. The correlation affects thereby the expected profit of the auctioneer. In the main result we conclude that a stronger positive correlation increases the expected profits of the auctioneer continuously, while a negative one decreases it.[1] This result is similar to the one by Laffont and Martimort (2000) who seek robustness of the optimal provision of public goods with respect to potential coalition among affected agents. In contrast to the finding in the uncorrelated environment (Bierbrauer and Netzer, 2016), the same expected profit cannot be materialized under the externality-freeness condition.

A study which combines ex post incentive compatibility with the externality-freeness condition is provided by Bierbrauer et al. (2017) for income taxation and bilateral trade environment. This more demanding incentive compatibility is able to ensure robustness against beliefs (Bergemann and Morris, 2005). The authors compare the approach by Mirrlees (1971) which satisfies externality-freeness with one that does not by Piketty (1993). The later one is based on perfect negative correlation which is an extreme case of the environment considered here. In an experimental study Bierbrauer et al. (2017) find that there is a significant larger de-

[1] Except for perfect positive correlation the expected payoff is strictly less than the one under public information. But in such an environment there is actually public information among the bidders.

viation from truthfully declaring the income by high income types in the mechanism by Piketty (1993) than in the one by Mirrlees (1971).

We conclude in the case of ex post incentive compatibility and externality-freeness that there are no differences between optimal auctions for correlated and uncorrelated valuation types. We show that externality-freeness implies ex post individual rationality. Both robustness requirements combined then demand a belief-free auction. But since the correlation only affects the beliefs of bidders and this channel cannot be used anymore in a belief-free auction, there cannot be any differences between correlated and uncorrelated environments.

In the next Section 5.2 we formally introduce the environment considered here, state the requirements on the optimal auction, and describe the auction by Crémer and McLean (1985) and Bartling and Netzer (2016) in the environment. In Section 5.3 we characterize the optimal auction under Bayesian incentive compatibility and state the relationship with the externality-robust auction. As the result crucially depends on the beliefs, we ensure in a second step also belief robustness and demand ex post incentive compatibility. We find that the optimal auction that satisfies robustness towards beliefs and social preferences is the same for correlated and uncorrelated valuation types. In the final Section 5.4 we conclude.

5.2 Framework

5.2.1 The Environment

We consider an environment where a single unit good is allocated by an auction to one of two ex ante symmetric bidders $i \in I = \{1, 2\}$. Both bidders have either a low or high valuations for the good, i.e. the private valuation type is $\theta_i \in \Theta_i = \{\theta^L, \theta^H\}$. A valuation type profile is $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 = \Theta$. The ex ante probability over valuation type profiles is $p : \Theta \to [0, 1]$.

Bidders' valuation types are assumed to be correlated. After bidders are informed about their own valuation type, they update their beliefs about the other bidder's valuation type to $p(\cdot|\theta_i) : \Theta_j \to [0, 1] \; \forall \, \theta_i \in \Theta_i$. The auctioneer is assumed to know the ex ante probability distribution of valuation type profiles. We define valuation types as positively correlated if $p(\theta^L|\theta^H) < p(\theta^L, \theta^L) + p(\theta^H, \theta^L)$ and as negatively correlated if $p(\theta^L|\theta^H) > p(\theta^L, \theta^L) + p(\theta^H, \theta^L)$.

A social choice function returns for each type profile an allocation. An allocation a specifies the fraction of the good each bidder receives (q_1, q_2) and the according transfers (t_1, t_2), i.e. $a = (q_1, q_2, t_1, t_2) \in A$. The amount to be allocated is restricted such that $q_1 + q_2 \in [0,1]$. The transfer can be any real number, i.e. $(t_1, t_2) \in \mathbb{R}^2$. An *environment* is therefore $\mathbb{V} = (\{1,2\}, \{\theta_1, \theta_2\}, p, A)$.

An auction Φ consists of a message profile $m = (m_1, m_2) \in M = M_1 \times M_2$ (bids) and an outcome function g that assigns an allocation to each message profile, i.e. $g : M \to A$. The outcome can be summarized by $g = (q^g, t^g)$ where $q^g = (q_1^g, q_2^g)$ is the allocation profile such that $q_i : M \to [0,1]$ for each $i \in \{1,2\}$. The according transfer profile $t^g = (t_1^g, t_2^g)$ consists of the individual transfer functions $t_i : M \to \mathbb{R}$ for each $i \in \{1,2\}$. We denote an auction by $\Phi = (M, g)$.

For each bidder a strategy is the mapping $s_i : \Theta_i \to M_i$. We denote a strategy profile by $s = (s_1, s_2)$. Given the *auction* Φ we assume that bidders material payoffs are quasi-linear, i.e.

$$\pi_i((s_i, s_j), \theta_i) = q_i(s_i, s_j)\ \theta_i + t_i(s_i, s_j). \tag{5.1}$$

5.2.2 Requirements on the Optimal Auction

The objective of the auctioneer is to maximize the expected profit $\pi_A = -\mathbb{E}[t_1 + t_2]$. As the valuation types of bidders are unknown to the auctioneer, we require that an optimal auction needs to satisfy incentive compatibility (IC)[2]. A second restriction is due to the individual rationality (IR). As we assume that bidders cannot be forced to participate in the auction, the expected material payoff for each valuation type should be at least as good as the outside option which we normalize to zero. For both (IC) and (IR) we denote the interim (Bayesian) versions with (BIC) and (BIR) and for the ex post versions with (EPIC) and (EPIR) respectively.

Crémer and McLean (1985) show that under interim individual rationality[3] there exists an auction that generates the same expected profit as if the auctioneer would know the valuation types of the participating bidders. From Chapter 4 we know that the auction design to achieve the result is sensitive to potential outcome-based social preferences among bidders. Hence, we require here that the optimal auction, in addition to the two standard constraints, satisfies an externality-freeness condition (EF) (Bierbrauer and Netzer, 2016).

The externality-freeness condition ensures robustness with respect to the unknown heterogeneity of social preferences among bidders. The condition states that the material payoff of one bidder (at least on the equilibrium path) is indepen-

[2]Incentive compatibility ensures that each valuation type has an (monetary) incentive to reveal truthful the valuation type.

[3]Interim individual rationality requires that the each bidder prefers to participate in the auction after she learns her own valuation type but before the valuation type of the other bidder is revealed. In contrast, the concept of ex post individual rationality requires that the each only prefers

dent of the strategy of the other bidder. Hence, if one bidder reveals her valuation type, the other bidder is unable to affect her material payoff. Another interpretation of the condition would be an insurance against the unknown valuation type of the other bidder. For any given social choice function Bierbrauer and Netzer (2016) provide a tool to construct a modified social choice function which satisfies the externality-freeness condition and generates the same expected profit (for details, see Proposition 2 in Bierbrauer and Netzer, 2016). The idea is based on a modification of the transfer function.

Definition 5.1. *An auction Φ satisfies the externality-freeness condition if and only if*

$$\pi((\theta_i, \theta^L), \theta_i) = \pi((\theta_i, \theta^H), \theta_i) \ \ \forall \theta_i \in \Theta_i \ \ \forall i \in I.$$

For the modification the authors assume that the valuation types are uncorrelated. But the valuation types we consider here are explicitly correlated. Hence, we analyze the full maximization problem stated in Crémer and McLean (1985) but require additionally the externality-freeness condition. We characterize the optimal auction for Bayesian and ex post incentive compatibility[4].

This characterization enables us to conduct several comparisons. One is with the result by Crémer and McLean (1985) and the effect of the social robustness on first best implementation. Further comparisons are possible with current insight about the effect of externality-freeness condition under the assumption of uncorrelated valuation types.

		social preferences: no	**social preferences**: yes
correlated valuations	no	Revenue Equivalence	Externality-Robust Auction
correlated valuations	yes	First Best Implementation	This Chapter

Table 5.1: Placement into the current literature

5.2.3 The Crémer&McLean and Bartling&Netzer Auctions

For the comparison later we describe first the auctions proposed by Bartling and Netzer (2016) (Φ^{BN}) and Crémer and McLean (1985) (Φ^{CM}) within the environment considered here.

Both auctions apply the revelation principle such that the message set is equivalent to the valuation type set. In our environment this means that $M^{BN} = M^{CM} =$

[4]Ex post incentive compatibility is sufficient to ensure robustness towards heterogeneity of beliefs among bidders (Bergemann and Morris, 2005)

$\{\theta^L, \theta^H\}$. We state the valuation type profile in the form $\theta = (\theta_i, \theta_j)$ such that it reports bidder i's valuation type first. Both auctions also provide an efficient allocation function, i.e.

$$q_i^{BN} = q_i^{CM} = \begin{cases} 0.5 & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ 1 & \text{if } \theta = (\theta^H, \theta^L), \\ 0.5 & \text{if } \theta = (\theta^H, \theta^H), \end{cases} \qquad \forall i \in \{1, 2\}. \tag{5.2}$$

The main difference between the two auctions is their transfer functions. The transfer function by Bartling and Netzer (2016) has the following form:

$$t_i^{BN}(m) = \mathbb{B}^{BN}(\theta_i) + \begin{cases} -0.5\theta^L & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ -\theta^H & \text{if } \theta = (\theta^H, \theta^L), \\ -0.5\theta^H & \text{if } \theta = (\theta^H, \theta^H), \end{cases} \qquad \forall i \in \{1, 2\} \tag{5.3}$$

where the bonus function depends solely on the own message m_i:

$$\mathbb{B}^{BN}(\theta_i) = \begin{cases} 0 & \text{if } \theta_i = \theta^L, \\ p(\theta^L)\; q_i(\theta^L, \theta^L)\; \Delta\theta & \text{if } \theta_i = \theta^H. \end{cases}$$

In summary, Φ^{BN} is a modified first price auction where both valuation types receive the expected gain from a deviation as a bonus to ensure Bayesian incentive compatibility. The bonus function $B^{BN}(\theta_i)$ is independent of the other bidder's valuation type θ_j which ensures the social robustness of the auction.

On the other hand, the transfer function by Crémer and McLean (1985) is

$$t_i^{CM}(\theta) = \mathbb{B}^{CM}(\theta_i) + \begin{cases} -0.5\theta^L & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ -\theta^L & \text{if } \theta = (\theta^H, \theta^L), \\ -0.5\theta^H & \text{if } \theta = (\theta^H, \theta^H). \end{cases} \qquad \forall i \in \{1, 2\} \tag{5.4}$$

Here, the bonus function depends solely on the other bidder's valuation type θ_j

(instead of solely on the own valuation type θ_i) and is given by

$$\mathbb{B}^{CM}(\theta_j) = \begin{cases} \frac{p(\theta^L|\theta^H)\, p(\theta^H|\theta^L)}{p(\theta^L|\theta^L)\, p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\, p(\theta^H|\theta^L)} \Delta\theta & \text{if } \theta_j = \theta^L \\ -\frac{p(\theta^L|\theta^L)\, p(\theta^L|\theta^H)}{p(\theta^L|\theta^L)\, p(\theta^H|\theta^H) - p(\theta^L|\theta^H)\, p(\theta^H|\theta^L)} \Delta\theta & \text{if } \theta_j = \theta^H. \end{cases}$$

The derivation of $\mathbb{B}^{CM}(\theta_j)$ is provided in the Appendix 4.A of Chapter 4. In contrast to Φ^{BN} , Φ^{CM} is a modified second price auction, not a first price auction, and the bonus function $B^{CM}(\theta_j)$ is independent of the own bid θ_i. The main differences between the two transfer functions t_i^{BN} and t_i^{CM} is the effect of the own strategy on the own and the other bidder's transfer.

In the next section we characterize the optimal auctions under the externality-freeness condition for two different equilibrium concepts. We label the auction Φ^{BIC} if Bayesian incentive compatibility is required and Φ^{EP} for the case of ex post incentive compatibility. The auction Φ^{BN} is the only one considering no correlation of valuation types. In contrast to Φ^{BN}, the auction Φ^{BIC} assumes that the valuation types are correlated but requests otherwise the same conditions.

As the auction Φ^{BIC} crucially depends on the conditional beliefs of the high valuation type, we require with Φ^{EP} robustness with respect to bidders' beliefs. The same equilibrium concept is used in Φ^{CM} which in contrast only demands interim individual rationality. At the beginning of Section 5.3 we provide with Lemma 2 a result that social robustness requires ex post individual rationality. Figure 5.1 displays the differences between the auction designs.

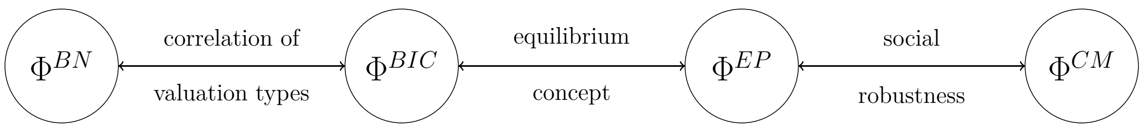


Figure 5.1: Comparisons between the auctions

5.3 Analysis

In this section we characterize the optimal auction design under Bayesian incentive compatibility (Φ^{BIC}) and ex post incentive compatibility (Φ^{EP}). This allows for the comparisons we introduced in Figure 5.1. We apply the revelation principle and focus on direct mechanism, i.e. we restrict the message set $M^g = \{\theta^L, \theta^H\}$ for all outcome functions. In Lemma 2 we state a general finding which reduces the set of constraints when we demand that the auction satisfies the externality-freeness condition. In general, we find that if we

Lemma 2. *Suppose that the auction Φ satisfies the externality-freeness condition.*
(1) Then the Bayesian individual rationality condition is satisfied if and only if both ex post individual rationality conditions are satisfied.
(2) Then for each valuation type there exists a unique ex post individual rationality condition.

Proof. Suppose the auction Φ satisfies the externality-freeness condition and w.l.o.g. an arbitrary bidder $i \in \{1, 2\}$. The externality-freeness condition requires that $\pi_i((\theta^L, \theta^L), \theta^L) = \pi_i((\theta^L, \theta^H), \theta^L)$ and $\pi_i((\theta^H, \theta^L), \theta^H) = \pi_i((\theta^H, \theta^H), \theta^H)$.
(1) For each type, the interim expected individual rationality is satisfied if and only if

$$\begin{aligned} &\mathbf{E}_{\theta_j}\left[\pi((\theta_i, \theta_j), \theta_i)\right] \geq 0 \text{ for each } \theta_i \in \{\theta^L, \theta^H\} \\ \Leftrightarrow\ & p(\theta^L|\theta_i)\pi((\theta_i, \theta^L), \theta_i) + p(\theta^H|\theta_i)\pi((\theta_i, \theta^L), \theta_i) \geq 0 \text{ for each } \theta_i \in \{\theta^L, \theta^H\} \end{aligned}$$

But since $\pi((\theta_i, \theta^L), \theta_i) = \pi((\theta_i, \theta^H), \theta_i)$ for each $i \in \{\theta^L, \theta^H\}$, every convex combination between the to ex post payoffs is equal. In particular,

$$\begin{aligned} \pi((\theta_i, \theta^L), \theta_i) &= p(\theta^L|\theta_i)\ \pi((\theta_i, \theta^L), \theta_i) + p(\theta^H|\theta_i)\ \pi((\theta_i, \theta^L), \theta_i) \\ &= \pi((\theta_i, \theta^H), \theta_i) \text{ for each } \theta_i \in \{\theta^L, \theta^H\}. \end{aligned}$$

Hence, Bayesian Interim individual rationality is satisfied if and only if both ex post individual rationality are satisfied.
(2) Moreover, because both ex post payoffs are equal, for each valuation type, the ex post individual rationality against one valuation type is satisfied if and only if the ex post individual rationality against the other valuation type is satisfied. This reduces the set of individual rationality which should be verified to one ex post individual rationality. □

The result is due to the formation of expected payoffs. Interim, the expected payoff of bidder i is the weighted sum of the ex post payoffs for the other bidder's valuation types. The weighting is given by the corresponding conditional probabilities of the other bidder's valuation types. Now, the externality-freeness condition demands that if bidder i reports truthful, all of these ex posts payoffs are equal. Hence any convex combination between the two ex post payoffs needs to be equal to them. This includes the possibility of the full support of both corners. And these corners coincide with the ex post payoffs against one certain valuation type. Hence, any interim expected payoff is equal to the ex post payoffs. And as they are equal, both ex post individual rationality constraints are equivalent. Therefore, only one of the need to be satisfied because then the other is automatically satisfied.

5.3.1 Optimal Auction for Bayesian Incentive Compatibility

We start with the weaker equilibrium concept that is also used by Bartling and Netzer (2016) for their auction Φ^{BN}. The authors require Bayesian incentive compatibility because their auction is a modified first price auction using the tool provided by Bierbrauer and Netzer (2016). As explained before the tool assumes uncorrelated valuation types. Hence, we are not able to simply apply it for the auction Φ^{CM} here. We therefore state the full maximization problem. In detail, the problem the auctioneers faces is

$$\max_{\{q^{BIC},\, t^{BIC}\}} -\mathbb{E}[t_1 + t_2] \qquad (objective)$$
$$s.t.\ (EF),\ (BIC),\ (EPIR)$$

Before we state the solution we reduce the constraints. Lemma 2 reduces the amount of individual rationality conditions because the externality-freeness condition implies the sufficiency of a unique ex post individual rationality. We state the full problem in Appendix 5.A. We show that

Proposition 5.1. *The optimal expected profit auction $\Phi^{BIC} = (M^{BIC}, (q^{BIC}, t^{BIC}))$ is a first price auction modified by a valuation type dependent bonus function $\mathbb{B}^{BIC}(\theta_i)$ and is given by*

$$q_i^{BIC}(\theta) = \begin{cases} 0 \text{ or } 0.5 & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ 1 & \text{if } \theta = (\theta^H, \theta^L), \\ 0.5 & \text{if } \theta = (\theta^H, \theta^H), \end{cases} \qquad \forall\, i \in \{1, 2\},$$

$$t_i^{BIC}(\theta) = \mathbb{B}^{BIC}(\theta_i) + \begin{cases} -q_i(\theta^L, \theta^L)\theta^L & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ -\theta^H & \text{if } \theta = (\theta^H, \theta^L), \\ -0.5\theta^H & \text{if } \theta = (\theta^H, \theta^H). \end{cases} \qquad \forall\, i \in \{1, 2\}.$$

The bonus function is given by

$$\mathbb{B}^{BIC}(\theta_i) = \begin{cases} 0 & \text{if } \theta_i = \theta^L, \\ p(\theta^L|\theta^H)\ q_i(\theta^L, \theta^L)\ \Delta\theta & \text{if } \theta_i = \theta^H. \end{cases}$$

Proof. See the derivation in Appendix 5.A. □

The allocation function in q_i^{BIC} is equal to q_i^{BN} by Bartling and Netzer (2016). It allocates the good to the bidder with the highest valuation. As the Bayesian incentive compatibility ensures that truthful revelation is preferred by the bidder, the allocation is efficient (material welfare maximizing) if $q_i^{BIC}(\theta^L, \theta^L) = 0.5$. Whether the allocation function assigns the good to one of the bidder when both valuations are low depends on the optimal reserve price. Proposition 5.2 discusses how the optimal reserve price depends on the valuation correlation. This incentive compatibility is guaranteed due to the modification of the transfer function by the bonus function.

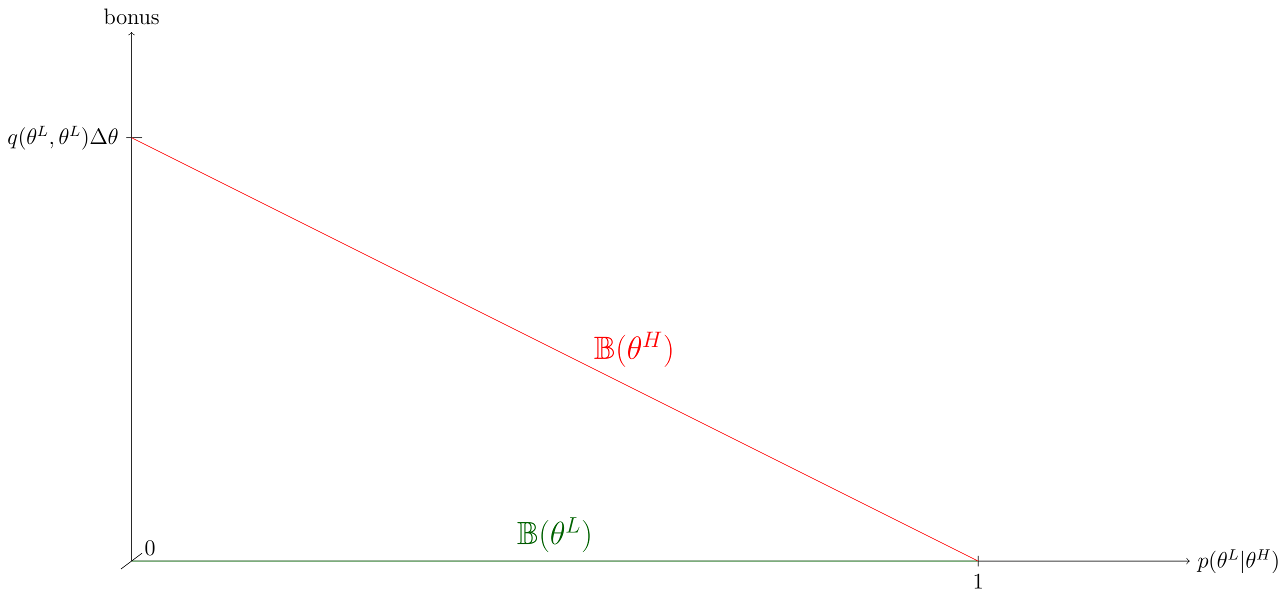


Figure 5.2: The dependence of the bonus function on the conditional probability

The role of the bonus function

In a standard first price auction the winning bidder simply pays her bid. In such a case a high valuation type prefers to deviate if the other bidder's valuation is low. Although it decreases the chance of winning, it increases the payoff conditional on winning from zero to the differences in valuations $\Delta\theta$. The high valuation type would be indifferent between both messages, if the other bidder bids a high valuation because in any case she would receive a zero payoff. Overall, a deviation is preferred in the first price auction for any $p(\theta^L|\theta^H)$.

The bonus function is designed such that the high valuation type expects the same material payoff from truthful revelation and the deviation. That is the difference in the valuation weighted by the conditional probability, which reflects her beliefs that the other bidder's valuation is indeed low, given the chance of winning in that case of 0.5. This bonus payment ensures that the high valuation type is indifferent between the truthful revelation of her valuation type and not reveling it. In addition, it is independent from the strategy played by the other bidder. On the other hand, it is never profitable for the low valuation type to deviate because the transfer $t_i(\theta^H)$ exceeds the valuation θ^L for all conditional probability $p(\theta^L|\theta^H)$.

In contrast to the uncorrelated environment considered by Bartling and Netzer (2016) the belief about the valuation type of the other bidder here is not independent of the own type. Hence the auctioneer needs to considering the updated beliefs of the valuation types.[5] Therefore the bonus function depends in our environment on the conditional beliefs. The differences between the ex ante and the interim beliefs about the other's bidder depend on the correlation of valuation types.

Comparison to the uncorrelated environment

In comparison to the uncorrelated case of Bartling and Netzer (2016) we find that

Theorem 5.1. *The expected profit from the auction increases (decreases) continuously in the intensity of positive (negative) correlations of valuation types.*

Proof. Consider the two auctions Φ^{BIC} and Φ^{BN}. Assume either correlated valuation types such that

$$p(\theta^L|\theta^H) - [p(\theta^L,\theta^L) + p(\theta^H,\theta^L)] \in [-1,1]\backslash\{0\} \tag{5.5}$$

where we define valuation types as positively correlated if $p(\theta^L|\theta^H) < p(\theta^L,\theta^L) + p(\theta^H,\theta^L)$ and as negatively correlated if $p(\theta^L|\theta^H) > p(\theta^L,\theta^L)+p(\theta^H,\theta^L)$. We denote valuation types as uncorrelated if and only if

$$p(\theta^L|\theta^H) - [p(\theta^L,\theta^L) + p(\theta^H,\theta^L)] = 0. \tag{5.6}$$

In general, the expected profit for the auctioneer is given by $-\mathbb{E}_\theta[t_1^g + t_2^g]$. For our

[5]The behavioral economic literature reports that experimental subjects do not update correctly (Tversky and Kahneman, 1973). Hence, it is not sure whether bidders are able to correctly apply the additional information.

two auctions with bonus $\mathbb{B}(\theta^H)$ this corresponds to

$$p(\theta^L,\theta^L)\theta^L + [p(\theta^H,\theta^H)+p(\theta^H,\theta^L)][\theta^H - \mathbb{B}(\theta^H)] + p(\theta^H,\theta^H)[\theta^H - 2\mathbb{B}(\theta^H)] \quad (5.7)$$

where $\mathbb{B}(\theta^H)$ is either

$$\mathbb{B}^{BIC}(\theta^H) = p(\theta^L|\theta^H)\frac{1}{2}\Delta\theta,$$
$$\mathbb{B}^{BN}(\theta^H) = \Big[p(\theta^L,\theta^L)+p(\theta^H,\theta^L)\Big]\frac{1}{2}\Delta\theta.$$

The difference in the expected profits between the two auctions is then

$$-\Big[p(\theta^L|\theta^H)-[p(\theta^L,\theta^L)+p(\theta^H,\theta^L)]\Big]\Big[p(\theta^L,\theta^H)+p(\theta^H,\theta^L)+2p(\theta^H,\theta^H)\Big]\frac{1}{2}\Delta\theta \quad (5.8)$$

The first derivative with respect to $p(\theta^L|\theta^H) - [p(\theta^L,\theta^L)+p(\theta^H,\theta^L)]$ is decreasing, i.e.

$$-[p(\theta^L,\theta^H)+p(\theta^H,\theta^L)+2p(\theta^H,\theta^H)]\frac{1}{2}\Delta\theta. \quad (5.9)$$

This implies that the more positive the correlation, the larger the expected profit of the auctioneer in comparison to no correlation. It decreases continuously in the intensity of the negative correlation. □

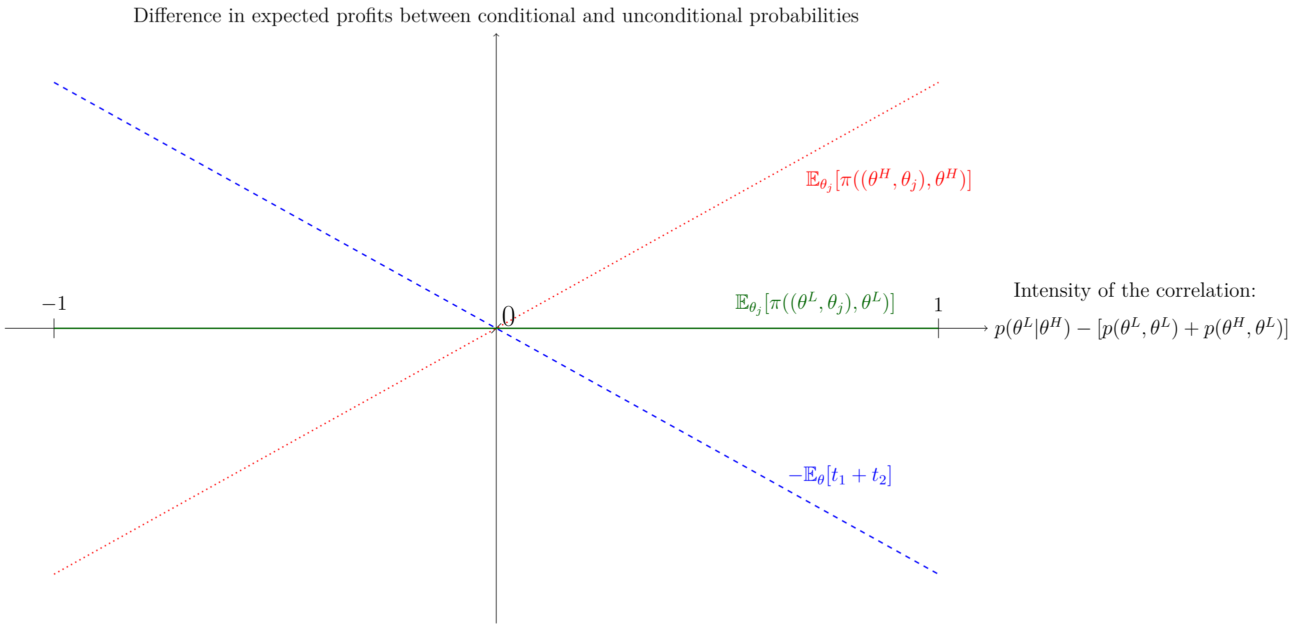


Figure 5.3: The effect of the correlation on expected payoffs and profits

In Figure 5.3 we illustrate the relationship of the correlation on the distribution of trade gains. Given that $q_i(\theta^L,\theta^L) = 0.5$ for both $i \in \{1,2\}$, the expected profit for the auctioneer of Φ^{BIC} $(-\mathbb{E}[t_1^B + t_2^B])$ is equal to

$$p(\theta^L,\theta^L)\theta^L + [p(\theta^L,\theta^H)+p(\theta^H,\theta^L)][\theta^H - \underbrace{\frac{1}{2}p(\theta^L|\theta^H)\Delta\theta}_{\mathbb{B}^{BIC}(\theta^H)}] + p(\theta^H,\theta^H)[\theta^H - \underbrace{p(\theta^L|\theta^H)\Delta\theta}_{2\times\mathbb{B}^{BIC}(\theta^H)}] \tag{5.10}$$

In equation (5.10) we see that the bonus function $\mathbb{B}^{BIC}(\theta^H)$ directly affects the expected profits of the auctioneer. This bonus function increases in the high valuation types' beliefs that the other bidder has a low valuation type, i.e.

$$\frac{\partial B^{BIC}(\theta^H)}{\partial p(\theta^L|\theta^H)} = \frac{1}{2}\Delta\theta > 0.$$

The stronger the positive correlation between the two types, the lower the resulting $p(\theta^L|\theta^H)$ because

$$p(\theta^L|\theta^H) = \frac{p(\theta^L,\theta^H)}{p(\theta^L,\theta^L)+p(\theta^L,\theta^H)}$$

decreases if $p(\theta^L,\theta^L)$ increases.

It is easy to verify that in the case of perfect positive correlation, i.e. $p(\theta^L|\theta^H) = 0$, the same expected profit as under symmetric information is achieved. But for any $p(\theta^L|\theta^H) > 0$ the auction materialized strictly less expected profit.

Hence, in comparison with the uncorrelated environment it is generally not possible to extract the same expected payoff under the externality-freeness condition in the less demanding equilibrium concept. Given selfish preferences, the expected profit is strictly less than under the Φ^{CM} if the correlation is not perfectly positive. If we in contrast assume non-selfish preferences, the comparison depends on the kind of social preferences and its distribution among bidders.

Optimal Reservation Price

We find that the influence of the correlation on $p(\theta^L|\theta^H)$ affects the optimal reserve price in the auction. Whether the low valuation type should be excluded from the auction depends on the belief that the high valuation type has about the probability that the other bidders' valuation is indeed low. We find that

Proposition 5.2. *The low valuation type is included, i.e.* $q_i(\theta^L,\theta^L) = 0.5$*, if and only if*

$$\frac{p(\theta^L,\theta^L)2\theta^L}{[2p(\theta^H,\theta^H)+p(\theta^H,\theta^L)+p(\theta^L,\theta^H)]\Delta\theta} \geq p(\theta^L|\theta^H) \tag{5.11}$$

and otherwise $q_i(\theta^L,\theta^L) = 0$ *for both* $i \in \{1,2\}$.

Proof. See Appendix 5.A. □

Hence, low valuation types are included when the probability that both bidders have a low valuation is sufficiently high. The threshold for that case depends on the conditional probability the high valuation type assigns to the other bidder's valuation types. This resembles the intensity of the correlation between the bidders valuation types. As with an increase of negative correlation, i.e. $p(\theta^L|\theta^H) \to 1$, the bonus $\mathbb{B}^{BIC}(\theta^H)$ increases, the likelihood that low valuation types are excluded increases.

5.3.2 Optimal Auction under Ex Post Incentive Compatibility

In the previous section we conclude that the optimal auction under Bayes incentive constraint requires that the auctioneer has perfect knowledge about the conditional probabilities the high valuation type assigns to the others bidder's valuations. Here, we seek additional robustness against such specific knowledge about beliefs (see for instance Bergemann and Morris, 2005, for a general criticism). A sufficient approach to overcome this specific knowledge problem is to require ex post incentive compatibility. Our main interest is in the effect of the additional constraint on the optimal auction design. We refer to the resulting auction as Φ^{EP}. We find that in the case of correlated valuation types the resulting expected profit for the auctioneer is lower.

Comparison to the uncorrelated environment

As the externality-freeness condition is the same as in the last subsection, we reduce again by Lemma 2 the set of ex post individual rationality conditions to a unique one for each valuation type. Ex post incentive compatibility requires intuitively that for each bidder each valuation type prefers to reveal her valuation type against each possible valuation type of the other bidder. For the participation as well as the strategy decisions beliefs no longer matter. But beliefs are the channel which differs for auction design in correlated and uncorrelated valuation types.

In general, we conclude therefore that

Proposition 5.3. *Under robustness conditions against uncertainty of social preferences and bidders' beliefs, the optimal auction mechanism is equivalent for correlated and uncorrelated valuation types.*

Proof. Suppose the auction (M, g) satisfies the externality-freeness condition and ex post incentive compatibility. Then by Lemma 2 we know that the externality-freeness condition implies that there exists a unique ex post individual rationality constraint for each valuation type. This implies that beliefs cannot affect the participation decision. Ex post incentive compatibility requires that reveling the own valuation type is optimal against each potential valuation type of the other bidder. This implies that the equilibrium strategy is independent of beliefs. In general, the participation and the optimal strategy are independent of bidders' beliefs. This belief-freeness implies that there are no difference between the optimal auction of correlated and uncorrelated valuation types. □

In other words, the auction designer is not able to use the additional information that bidders have about each other in correlated environments under this set of constraints to increase the expected profit.

The optimal auction

We provide a full characterization of the optimal auction under ex post incentive compatibility and the externality-freeness condition. This characterization enables to us to compare the expected profit of the auctioneer to the one under Bayesian incentive compatibility. Our analysis is comparable to work by Bierbrauer et al. (2017) who investigate socially robust implementation of a social choice function in two different environments: bilateral trade and redistributive income taxation.

Here, we again restrict our analysis on a direct auction mechanism $(\{\theta^L, \theta^H\}, g^{EPIC})$, and the problem the auctioneers faces is

$$\max_{\{q^{EP},\, t^{EP}\}} -\mathbb{E}[t_1 + t_2] \qquad (objective)$$

$$s.t.\ (EF),\ (EPIC),\ (EPIR)$$

Proposition 5.4. *The optimal expected profit auction* $\Phi^{EP} = (M^{EP}, (q^{EP}, t^{EP}))$ *is a first price auction with a valuation type dependent bonus function* $\mathbb{B}^{EP}(m_i)$ *and given by*

$$q_i^{EP}(\theta) = \begin{cases} 0 \text{ or } 0.5 & \text{if } \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \theta = (\theta^L, \theta^H), \\ 1 & \text{if } \theta = (\theta^H, \theta^L), \\ 0.5 & \text{if } \theta = (\theta^H, \theta^H), \end{cases}$$

$$t_i^{EP}(\theta) = \mathbb{B}^{EP}(\theta_i) + \begin{cases} -q_i(\theta^L, \theta^L)\theta^L & \textit{if } \theta = (\theta^L, \theta^L), \\ 0 & \textit{if } \theta = (\theta^L, \theta^H), \\ -\theta^H & \textit{if } \theta = (\theta^H, \theta^L), \\ -0.5\theta^H & \textit{if } \theta = (\theta^H, \theta^H). \end{cases}$$

where the bonus function is

$$\mathbb{B}^{EP}(\theta_i) = \begin{cases} 0 & \textit{if } \theta_i = \theta^L \\ q(\theta^L, \theta^L)\ \Delta\theta & \textit{if } \theta_i = \theta^H \end{cases}$$

Proof. See the derivation in appendix 5.B. □

The expected profit for $q_i(\theta^L, \theta^L) = 0.5$ of the auction is

$$-\mathbb{E}[t_1 + t_2] = [p(\theta^L, \theta^L) + p(\theta^H, \theta^H)]\ \theta^L + [p(\theta^H, \theta^L) + p(\theta^L, \theta^H)]\ [0.5\theta^H + \theta^L] \tag{5.12}$$

As there has to be paid a strictly positive information rent to the high valuation type in the presence of a low valuation type, the expected revenue is strictly less than the under first best. In addition, the result is independent of the belief structure. In contrast to the case of Bayesian incentive compatibility, the allocation rule does not depend on the intensity of correlation anymore.

Proposition 5.5. *The low valuation type is not excluded, i.e.* $q(\theta^L, \theta^L) = 0.5$*, if and only if*

$$\frac{p(\theta^L, \theta^L)2\theta^L}{[2p(\theta^H, \theta^H) + p(\theta^H, \theta^L) + p(\theta^L, \theta^H)]\Delta\theta} \geq 1 \tag{5.13}$$

In comparison to Φ^{BIC} *the exclusion of low valuation types is more likely.*

Proof. See appendix. □

Here, it simply depends on the relative frequency whether the low valuation type occurs as in the case of correlated valuations types. In comparison with Φ^{BIC} it is more likely that the low valuation type is excluded to increase the expected profit which leads to an additional inefficiency in the case of actually two low valuation types. As the bonus function is larger than in the case of Bayes incentive compatibility, the exclusion of low valuation types leads to an increase of the expected profit due to the redundance of the bonus function.

Proposition 5.6. *The expected profit of the auctioneer in* Φ^{BIC} *depends on the correlation between valuation types and lies in the interval of the realizable expected profits by* Φ^{CM} *and* Φ^{EP}*.*

Figure 5.4 illustrates how the expected profits of the auctioneer in the three auctions depend on the correlation under the assumption of selfish preferences and that the valuations are correlated, i.e. $p(\theta^L|\theta^H) \neq p(\theta^L)$. The expected profits are highest in Φ^{CM} and lowest in Φ^{EP} and both are independent of the intensity of the intensity of the correlation between valuation types. In contrast, the profitability of Φ^{BIC} depends on the intensity because the bonus function is conditional on the interim beliefs of the high valuation type. As Φ^{EP} is robust towards all possible interim beliefs, the bonus function is equal to the highest possible bonus paid to the high valuation type in Φ^{BIC}. Therefore, the expected profitability of Φ^{BIC} is larger than the one of Φ^{EPIC}.

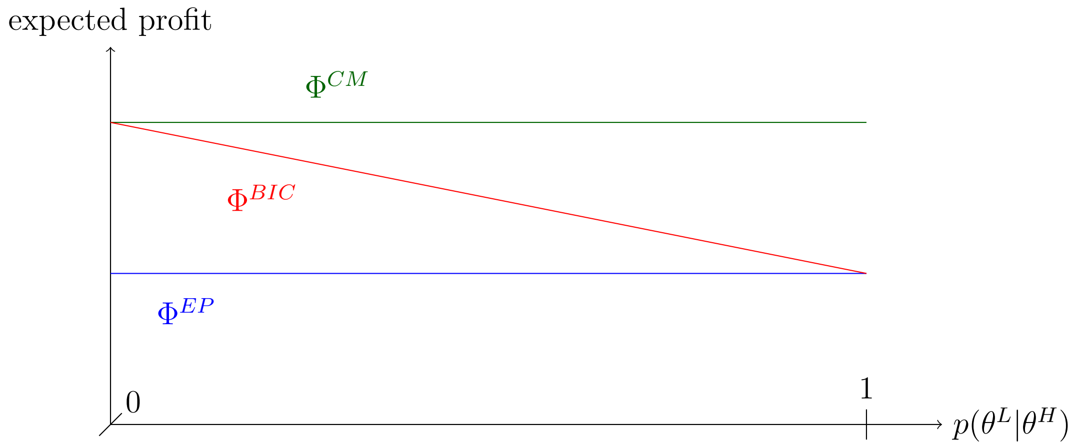


Figure 5.4: Expected profits for the auctioneer

5.4 Conclusion

The previous Chapter 4 shows that the first best implementation in non-independent environments as proposed by Crémer and McLean (1985) is not robust towards the possibility of outcome-based social preferences. A sufficient solution to this problem we investigate here is the externality-freeness condition by Bierbrauer and Netzer (2016). We study the effect of this additional condition under two equilibrium concepts: Bayes Nash and ex post equilibrium.

If we demand Bayesian incentive compatibility such that bidders reveal truthful their valuations in expectation, we show that the expected profit and the payoff of a high valuation type depend on the correlation between valuation types. In contrast to the standard literature under selfish preferences, we show that there exists a continuous relationship between the valuation correlation and the expected profit of the auctioneer.

In the case of ex post incentive compatibility we show that the valuation correlation does not affect the optimal design of an auction. The additional robustness implies that all optimal strategies are belief-free and therefore the same for any kind of correlation including the uncorrelated one.

Overall, we conclude that there is either a continuous relationship between correlated and uncorrelated valuations in the case of Bayesian incentive compatibility or even no difference if incentive compatibility is required ex post. This is in contrast to the standard conclusion by Crémer and McLean (1985, 1988) which shows that there exists a discontinuity between correlated and uncorrelated valuation type distribution.

5.A Expected profit maximization under BIC

Proposition 5.1

We assume in the environment that the two agents are ex ante symmetric. Therefore we show the optimal auction assuming an arbitrary agent $i \in \{1, 2\}$. The types are indicates by L for the valuation θ^L and, respectively, H for θ^H. The full problem the auctioneer faces is:

$$\max_{\{q^{BIC}, t^{BIC}\}} -\mathbb{E}_\theta[t_1(\theta_1, \theta_2) + t_2(\theta_2, \theta_1)] \tag{$objective$}$$

subject to

$$q_i(\theta^L, \theta^L)\theta^L + t_i(\theta^L, \theta^L) = q_i(\theta^L, \theta^H)\theta^L + t_i(\theta^L, \theta^H) \tag{EF_L}$$
$$q_i(\theta^H, \theta^L)\theta^H + t_i(\theta^H, \theta^L) = q_i(\theta^H, \theta^H)\theta^H + t_i(\theta^H, \theta^H) \tag{EF_H}$$

$$\begin{aligned} & p_i(\theta^L|\theta^L)\big[q_i(\theta^L, \theta^L)\theta^L + t_i(\theta^L, \theta^L)\big] + p(\theta^H|\theta^L)\big[q_i(\theta^L, \theta^H)\theta^L + t_i(\theta^L, \theta^H)\big] \\ \geq\ & p(\theta^L|\theta^L)\big[q_i(\theta^H, \theta^L)\theta^L + t_i(\theta^H, \theta^L)\big] + p(\theta^H|\theta^L)\big[q_i(\theta^H, \theta^H)\theta^L + t_i(\theta^H, \theta^H)\big] \end{aligned} \tag{BIC_L}$$

$$\begin{aligned} & p(\theta^L|\theta^H)\big[q_i(\theta^H, \theta^L)\theta^H + t_i(\theta^H, \theta^L)\big] + p(\theta^H|\theta^H)\big[q_i(\theta^H, \theta^H)\theta^H + t_i(\theta^H, \theta^H)\big] \\ \geq\ & p(\theta^L|\theta^H)\big[q_i(\theta^L, \theta^L)\theta^H + t_i(\theta^L, \theta^L)\big] + p(\theta^H|\theta^H)\big[q_i(\theta^L, \theta^H)\theta^H + t_i(\theta^L, \theta^H)\big] \end{aligned} \tag{BIC_H}$$

$$t_i(\theta^L, \theta^H) \geq -q_i(\theta^L, \theta^H)\theta^L \tag{$EPIR_L$}$$
$$t_i(\theta^H, \theta^L) \geq -q_i(\theta^H, \theta^L)\theta^H \tag{$EPIR_H$}$$

From the externality freeness conditions (EF_L) and (EF_H) we know how $t(\theta^L, \theta^L)$ and $t(\theta^H, \theta^H)$ are a function of $t(\theta^L, \theta^H)$ or $t(\theta^L, \theta^H)$. Hence the maximum possible expected profit is if both $(EPIR)$ are binding, i.e.

$$t_i(\theta^L, \theta^H) = -q_i(\theta^L, \theta^H)\theta^L \tag{$EPIR_L$}$$
$$t_i(\theta^H, \theta^L) = -q_i(\theta^H, \theta^L)\theta^H \tag{$EPIR_H$}$$

While (BIC_L) is slacked, one can show that under both binding $(EPIR)$ the (BIC_H) is not satisfied, i.e.

$$\begin{aligned} t_i(\theta^H,\theta^L) = & -q_i(\theta^H,\theta^L)\theta^H \geq -q_i(\theta^H,\theta^L)\theta^H \\ & + \underbrace{\left[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\right]\times(\theta^H-\theta^L)}_{>0}. \end{aligned}$$

For second best profit maximization (BIC_H) has to bind:

$$\begin{aligned} t_i(\theta^H,\theta^L) = & -q_i(\theta^H,\theta^L)\theta^H \\ & + \underbrace{\left[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\right]\times(\theta^H-\theta^L)}_{=\,Information\ Rent} \end{aligned}$$

$(EPIR_H)$ is obviously satisfied if (BIC_H) binds. (BIC_L) is satisfied if and only if

$$\begin{aligned} t_i(\theta^L,\theta^H) \geq & -q_i(\theta^L,\theta^H)\theta^L + p(\theta^L|\theta^L)q_i(\theta^H,\theta^L)\theta^L \\ & + p(\theta^H|\theta^L)\left[q_i(\theta^H,\theta^L)\theta^H + q_i(\theta^H,\theta^H)(\theta^L-\theta^H)\right] + t_i(\theta^H,\theta^L) \\ \Leftrightarrow\quad 0 \geq & \big(p(\theta^L|\theta^L)[q_i(\theta^H,\theta^H) - q_i(\theta^H,\theta^L)] + [q_i(\theta^L,\theta^H) - q_i(\theta^H,\theta^H)] \\ & + p(\theta^L|\theta^H)[q_i(\theta^L,\theta^L) - q_i(\theta^L,\theta^H)]\big)\times[\theta^H-\theta^L] \end{aligned}$$

With other words it is sufficient for (BIC_L) that $q_i(\theta^H,\theta^L) \geq q_i(\theta^H,\theta^H) \geq q_i(\theta^L,\theta^H)$ and $q_i(\theta^L,\theta^L) \geq q_i(\theta^L,\theta^H)$. We later control whether the allocation function satisfies the condition in the optimum. Hence the resulting transfer function is

$$\begin{aligned} t_i(\theta^L,\theta^L) &= -q_i(\theta^L,\theta^L)\theta^L \\ t_i(\theta^L,\theta^H) &= -q_i(\theta^L,\theta^H)\theta^L \\ t_i(\theta^H,\theta^L) &= -q_i(\theta^H,\theta^L)\theta^H + \left[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\right]\times(\theta^H-\theta^L) \\ t_i(\theta^H,\theta^H) &= -q_i(\theta^H,\theta^H)\theta^H + \left[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\right]\times(\theta^H-\theta^L) \end{aligned}$$

The expected profit $-\mathbb{E}[t_1+t_2]$ is

$$\begin{aligned} & p(\theta^L,\theta^L)[-2t_i(\theta^L,\theta^L)] + [p(\theta^L,\theta^H) + p(\theta^H,\theta^L)][-t_i(\theta^L,\theta^H) - t_i(\theta^H,\theta^L)] \\ & + p(\theta^L,\theta^L)[-2t_i(\theta^H,\theta^H)] \end{aligned} \tag{5.14}$$

Plug in:

$$
\begin{aligned}
&p(\theta^L,\theta^L)[2q_i(\theta^L,\theta^L)\theta^L] + [p(\theta^L,\theta^H)+p(\theta^H,\theta^L)][q_i(\theta^L,\theta^H)\theta^L + q_i(\theta^H,\theta^L)\theta^H \\
&-\big[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\big]\times(\theta^H-\theta^L)] \\
&+p(\theta^H,\theta^H)2[q_i(\theta^H,\theta^H)\theta^H - \big[p(\theta^L|\theta^H)\times q_i(\theta^L,\theta^L) + p(\theta^H|\theta^H)\times q_i(\theta^L,\theta^H)\big]\times(\theta^H-\theta^L)]
\end{aligned}
$$

Taking the first derivative with respect to the allocation function one finds that

$$
\begin{aligned}
q_i(\theta^H,\theta^H) &= 0.5 \\
q_i(\theta^L,\theta^H) &= 0 \\
q_i(\theta^H,\theta^L) &= 1
\end{aligned}
$$

while $q_i(\theta^L,\theta^L)\in\{0,0.5\}$ depends on the probability $p(\theta^L,\theta^L)$.
The transfer function is

$$
t_i^{BIC}(\theta) = \mathbb{B}^{BIC}(\theta_i) + \begin{cases} -q_i(\theta^L,\theta^L)\theta^L & \text{if } \theta=(\theta^L,\theta^L), \\ 0 & \text{if } \theta=(\theta^L,\theta^H), \\ -\theta^H & \text{if } \theta=(\theta^H,\theta^L), \\ -0.5\theta^H & \text{if } \theta=(\theta^H,\theta^H). \end{cases} \qquad \forall\, i\in\{1,2\}.
$$

with

$$
\mathbb{B}^{BIC}(\theta_i) = \begin{cases} 0 & \text{if } \theta_i=\theta^L, \\ p(\theta^L|\theta^H)\; q_i(\theta^L,\theta^L)\;\Delta\theta & \text{if } \theta_i=\theta^H. \end{cases}
$$

Proposition 5.2

To set $q_i(\theta^L,\theta^L)=0.5$ is optimal if and only if the expected profits $-\mathbb{E}[t_1+t_2]$ increase in $q_i(\theta^L,\theta^L)$. We find that the first derive with respect to $q_i(\theta^L,\theta^L)$ is greater than zero if and only if

$$
\begin{aligned}
& p(\theta^L,\theta^L)2\theta^L + [p(\theta^L,\theta^H)+p(\theta^H,\theta^L)][-p(\theta^L|\theta^H)\Delta\theta] + p(\theta^H,\theta^H)[-2p(\theta^L|\theta^H)\Delta\theta] \geq 0 \\
\Leftrightarrow\ & p(\theta^L,\theta^L)2\theta^L \geq [p(\theta^L,\theta^H)+p(\theta^H,\theta^L)][-p(\theta^L|\theta^H)\Delta\theta] + p(\theta^H,\theta^H)[-p(\theta^L|\theta^H)2\Delta\theta] \\
\Leftrightarrow\ & \frac{p(\theta^L,\theta^L)2\theta^L}{[p(\theta^L,\theta^H)+p(\theta^H,\theta^L)][-\Delta\theta] + p(\theta^H,\theta^H)[-2\Delta\theta]} \geq p(\theta^L|\theta^H) \qquad (5.15)
\end{aligned}
$$

As $q_i(\cdot)$ is upper bounded by 1 the highest possible $q_i(\theta^L,\theta^L)=0.5$. The lower bound is $q_i(\theta^L,\theta^L)=0$ in the case inequality (5.15) is not satisfied. Note that in either case the sufficient condition for (BIC_L) is satisfied.

5.B Expected profit maximization under EPIC

Proposition 5.4

We assume in the environment that the two agents are ex ante symmetric. Therefore we show the optimal auction assuming an arbitrary agent $i \in \{1, 2\}$. The types are indicates by L for the valuation θ^L and, respectively, H for θ^H. The full problem the auctioneer faces is:

$$\max_{\{q^{EP}, t^{EP}\}} -\mathbb{E}[t_1 + t_2] \qquad (objective)$$

subject to

$$q_i(\theta^L, \theta^L)\theta^L + t_i(\theta^L, \theta^L) = q_i(\theta^L, \theta^H)\theta^L + t_i(\theta^L, \theta^H) \qquad (EF_L)$$
$$q_i(\theta^H, \theta^H)\theta^H + t_i(\theta^H, \theta^H) = q_i(\theta^H, \theta^L)\theta^H + t_i(\theta^H, \theta^L) \qquad (EF_H)$$

$$q_i(\theta^L, \theta^L)\theta^L + t_i(\theta^L, \theta^L) \geq q_i(\theta^H, \theta^L)\theta^L + t_i(\theta^H, \theta^L) \qquad (EPIC_{LL})$$
$$q_i(\theta^L, \theta^H)\theta^L + t_i(\theta^L, \theta^H) \geq q_i(\theta^H, \theta^H)\theta^L + t_i(\theta^H, \theta^H) \qquad (EPIC_{LH})$$
$$q_i(\theta^H, \theta^L)\theta^H + t_i(\theta^H, \theta^L) \geq q_i(\theta^L, \theta^L)\theta^H + t_i(\theta^L, \theta^L) \qquad (EPIC_{HL})$$
$$q_i(\theta^H, \theta^H)\theta^H + t_i(\theta^H, \theta^H) \geq q_i(\theta^L, \theta^H)\theta^H + t_i(\theta^L, \theta^H) \qquad (EPIC_{HH})$$

$$t_i(\theta^L, \theta^H) \geq -q_i(\theta^L, \theta^H)\theta^L \qquad (EPIR_L)$$
$$t_i(\theta^H, \theta^L) \geq -q_i(\theta^H, \theta^L)\theta^H \qquad (EPIR_H)$$

In a first step we check whether the externality freeness condition is compatible with the ex post incentive compatibility. For that purpose we write the first six conditions in terms of $t(\theta^L, \theta^L)$ and $t(\theta^H, \theta^H)$:

$$t(\theta^L, \theta^L) = [q(\theta^L, \theta^H) - q(\theta^L, \theta^L)]\theta^L + t(\theta^L, \theta^H)$$
$$t(\theta^H, \theta^H) = [q(\theta^H, \theta^L) - q(\theta^H, \theta^H)]\theta^H + t(\theta^H, \theta^L)$$
$$t(\theta^L, \theta^L) \geq [q(\theta^H, \theta^L) - q(\theta^L, \theta^L)]\theta^L + t(\theta^H, \theta^L)$$
$$t(\theta^H, \theta^H) \leq [q(\theta^L, \theta^H) - q(\theta^H, \theta^H)]\theta^L + t(\theta^L, \theta^H)$$
$$t(\theta^L, \theta^L) \leq [q(\theta^H, \theta^L) - q(\theta^L, \theta^L)]\theta^H + t(\theta^H, \theta^L)$$
$$t(\theta^H, \theta^H) \geq [q(\theta^L, \theta^H) - q(\theta^H, \theta^H)]\theta^H + t(\theta^L, \theta^H)$$

The six conditions are satisfies if and only if the following four conditions hold

$$t(\theta^L,\theta^H) \geq [q(\theta^H,\theta^L) - q(\theta^L,\theta^H)]\theta^L + t(\theta^H,\theta^L) \tag{5.16}$$
$$t(\theta^L,\theta^H) \geq [q(\theta^H,\theta^L) - q(\theta^H,\theta^H)]\theta^H - [q(\theta^L,\theta^H) - q(\theta^H,\theta^H)]\theta^L + t(\theta^H,\theta^L) \tag{5.17}$$
$$t(\theta^L,\theta^H) \leq [q(\theta^H,\theta^L) - q(\theta^L,\theta^L)]\theta^H - [q(\theta^L,\theta^H) - q(\theta^L,\theta^L)]\theta^L + t(\theta^H,\theta^L) \tag{5.18}$$
$$t(\theta^L,\theta^H) \leq [q(\theta^H,\theta^L) - q(\theta^L,\theta^H)]\theta^H + t(\theta^H,\theta^L) \tag{5.19}$$

Note that (5.18) is not satisfied under binding individual rationality

$$t_i(\theta^L,\theta^H) = -q_i(\theta^L,\theta^H)\theta^L \tag{5.20}$$
$$t_i(\theta^H,\theta^L) = -q_i(\theta^H,\theta^L)\theta^H \tag{5.21}$$

while (5.19) is not satisfied if $q_i(\theta^L,\theta^H) > 0$ which is not true in the optimum as we see later. The two conditions for the low type are slacked in that case. Hence, let (5.18) bind. The resulting transfer is

$$t(\theta^H,\theta^L) = -q_i(\theta^H,\theta^L)\theta^H + q_i(\theta^L,\theta^L)[\theta^H - \theta^L]$$

and via the externality freeness condition we know that

$$t(\theta^H,\theta^H) = -q_i(\theta^H,\theta^H)\theta^H + q_i(\theta^L,\theta^L)[\theta^H - \theta^L].$$

For the low valuation type the transfer rules are

$$t_i(\theta^L,\theta^L) = -q_i(\theta^L,\theta^L)\theta^L$$
$$t_i(\theta^L,\theta^H) = -q_i(\theta^L,\theta^H)\theta^L.$$

The expected revenue for the auctioneer $-\mathbb{E}[t_1 + t_2]$ is

$$\begin{aligned}
&p(\theta^L,\theta^L)\big[2q_i(\theta^L,\theta^L)\theta^L\big]\\
&+ [p(\theta^L,\theta^H) + p(\theta^H,\theta^L)]\big[q_i(\theta^L,\theta^H)\theta^L + q_i(\theta^H,\theta^L)\theta^H - q_i(\theta^L,\theta^L)[\theta^H - \theta^L]\big]\\
&+ p(\theta^H,\theta^H)\big[2q_i(\theta^H,\theta^H)\theta^H - 2q_i(\theta^L,\theta^L)[\theta^H - \theta^L]\big]
\end{aligned}$$

Taking the first derivative with respect to the allocation rules, we receive

$$q_i(\theta) = \begin{cases} 0/0.5 & if\ \theta = (\theta^L, \theta^L) \\ 0 & if\ \theta = (\theta^L, \theta^H) \\ 1 & if\ \theta = (\theta^H, \theta^L) \\ 0.5 & if\ \theta = (\theta^H, \theta^H) \end{cases} \ for\ \ i \ \in \{1, 2\}.$$

The transfer function is

$$t_i^{EP}(\theta) = \mathbb{B}^{EP}(\theta_i) + \begin{cases} -q_i(\theta^L, \theta^L)\theta^L & \text{if } \ \theta = (\theta^L, \theta^L), \\ 0 & \text{if } \ \theta = (\theta^L, \theta^H), \\ -\theta^H & \text{if } \ \theta = (\theta^H, \theta^L), \\ -0.5\theta^H & \text{if } \ \theta = (\theta^H, \theta^H). \end{cases}$$

where the bonus function is

$$\mathbb{B}^{EP}(\theta_i) = \begin{cases} 0 & \text{if } \ \theta_i = \theta^L \\ q(\theta^L, \theta^L)\ \Delta\theta & \text{if } \ \theta_i = \theta^H \end{cases}$$

Proposition 5.5

To set $q_i(\theta^L, \theta^L) = 0.5$ is optimal if and only if the expected profits $-\mathbb{E}[t_1 + t_2]$ increase in $q_i(\theta^L, \theta^L)$. We find that the first derive with respect to $q_i(\theta^L, \theta^L)$ is greater than zero if and only if

$$\begin{aligned} & p(\theta^L, \theta^L)2\theta^L + [p(\theta^L, \theta^H) + p(\theta^H, \theta^L)][-\Delta\theta] + p(\theta^H, \theta^H)[-2\Delta\theta] \geq 0 \\ \Leftrightarrow \quad & p(\theta^L, \theta^L)2\theta^L \geq [p(\theta^L, \theta^H) + p(\theta^H, \theta^L)][-\Delta\theta] + p(\theta^H, \theta^H)[-2\Delta\theta] \\ \Leftrightarrow \quad & \frac{p(\theta^L, \theta^L)2\theta^L}{[p(\theta^L, \theta^H) + p(\theta^H, \theta^L)][-\Delta\theta] + p(\theta^H, \theta^H)[-2\Delta\theta]} \geq 1. \end{aligned} \tag{5.22}$$

In comparison to 5.15 we find that it the inequality 5.22 is satisfied only for a subset of parameter combinations and, hence, it is less likely that low valuation types are included. The reason that the bonus paid to the high valuation type is grater for all $p(\theta^L|\theta^H) \in [0, 1]$ in Φ^{EP} than in Φ^{BIC}.

Bibliography

Abeler, J., Nosenzo, D., and Raymond, C. (2016). Preferences for truth-telling. *IZA Discussion Paper*.

Akerlof, G. (1970). The market for "lemons": Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500.

Andreoni, J., Che, Y.-K., and Kim, J. (2007). Asymmetric information about rivals' types in standard auctions: An experiment. *Games and Economic Behavior*, 59(2):240–259.

Attiyeh, G., Franciosi, R., and Isaac, R. M. (2000). Experiments with the pivot process for providing public goods. *Public choice*, 102(1-2):93–112.

Bartling, B., Gesche, T., and Netzer, N. (2017). Does the absence of human sellers bias bidding behavior in auction experiments? *Journal of the Economic Science Association*, 3(1):44–61.

Bartling, B. and Netzer, N. (2016). An externality-robust auction: Theory and experimental evidence. *Games and Economic Behavior*, 97:186–204.

Benndorf, V., Rau, H. A., and Sölch, C. (2014). Minimizing learning behavior in experiments with repeated real-effort tasks. *Unpublished manuscript*.

Bergemann, D., Brooks, B., and Morris, S. (2016). Informationally robust optimal auction design. *Cowles Foundation Discussion Paper*, 2065.

Bergemann, D. and Morris, S. (2005). Robust mechanism design. *Econometrica*, 73(6):1771–1813.

Bierbrauer, F. and Netzer, N. (2016). Mechanism design and intentions. *Journal of Economic Theory*, 163:557–603.

Bierbrauer, F., Ockenfels, A., Pollak, A., and Rückert, D. (2017). Robust mechanism design and social preferences. *Journal of Public Economics*, 149:59–80.

Blewett, L. A., Ward, A., and Beebe, T. J. (2006). How much health insurance is enough? revisiting the concept of underinsurance. *Medical Care Research and Review*, 63(6):663–700.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, 115(506):1054–1076.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chatterjee, K. (1982). Incentive compatibility in bargaining under uncertainty. *The Quarterly Journal of Economics*, 97(4):717–726.

Chen, Y. (2008). Incentive-compatible mechanisms for pure public goods: A survey of experimental research. *Handbook of experimental economics results*, 1:625–643.

Chen, Y. and Chen, D. (2013). The review and analysis of compulsory insurance. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 4(1):6–17.

Chiappori, P.-A. and Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78.

Cohen, A. and Dehejia, R. (2004). The effect of automobile insurance and accident liability laws on traffic fatalities. *The Journal of Law and Economics*, 47(2):357–393.

Cohen, A. and Siegelman, P. (2010). Testing for adverse selection in insurance markets. *Journal of Risk and Insurance*, 77(1):39–84.

Coppinger, V. M., Smith, V. L., and Titus, J. A. (1980). Incentives and behavior in english, dutch and sealed-bid auctions. *Economic inquiry*, 18(1):1–22.

Cox, J. C., Roberson, B., and Smith, V. L. (1982). Theory and behavior of single object auctions. *Research in experimental economics*, 2(1):1–43.

Crémer, J. and McLean, R. P. (1985). Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53(2):345–361.

Crémer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257.

Crosetto, P. and Filippin, A. (2013). The bomb risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.

Cutler, D. M. and Zeckhauser, R. J. (1998). Adverse selection in health insurance. In *Forum for Health Economics and Policy*, volume 1.

Damgaard, M. T. and Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157:15–26.

Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise–a theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, 47(2):268–298.

Çelen, B., Schotter, A., and Blanco, M. (2017). On blame and reciprocity: Theory and experiments. *Journal of Economic Theory*, 169:62 – 92.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *mimeo.*

Falk, A., Fehr, E., and Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, 41(1):20–26.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

Fang, H., Keane, M. P., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy*, 116(2):303–350.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly journal of Economics*, pages 817–868.

Findling, M. A. and Germano, B. J. (1988). Underinsurance in indiana: An illusion of coverage. *Ind. L. Rev.*, 21:205.

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

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

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters*, 71(3):397–404.

Friesen, L. and Gangadharan, L. (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, 117(3):624 – 626.

Friesen, L. and Gangadharan, L. (2013). Designing self-reporting regimes to encourage truth telling: An experimental study. *Journal of Economic Behavior and Organization*, 94:90–102.

Fukukawa, K. (2002). Developing a framework for ethicallyquestionable behavior in consumption. *Journal of Business Ethics*, 41(1):99–119.

Furnival, G. M. and Wilson, R. W. (1974). Regressions by leaps and bounds. *Technometrics*, 16(4):499–511.

Gabaldón, I. M., Hernández, F. J. V., and Watt, R. (2014). The effect of contract type on insurance fraud. *Journal of Insurance Regulation*, 33:197.

Gale, D. and Hellwig, M. (1985). Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies*, 52(4):647–663.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society*, pages 587–601.

Gizatulina, A. and Hellwig, M. (2017). The generic possibility of full surplus extraction in models with large type spaces. *Journal of Economic Theory*, 170:385 – 416.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, 95(1):384–394.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–453.

Gollier, C. (1987). Pareto-optimal risk sharing with fixed costs per claim. *Scandinavian Actuarial Journal*, 1987(1-2):62–73.

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

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3(4):367 – 388.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–82.

Harsanyi, J. C. (1967). Games with incomplete information played by bayesian players, i–iii part i. the basic model. *Management science*, 14(3):159–182.

Harsanyi, J. C. (1968a). Games with incomplete information played by bayesian players part ii. bayesian equilibrium points. *Management Science*, 14(5):320–334.

Harsanyi, J. C. (1968b). Games with incomplete information played by bayesian-players, part iii. the basic probability distribution of the game. *Management Science*, 14(7):486–502.

Hölmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91.

Hurwicz, L. (1972). On informationally decentralized systems. *Decision and Organization: A Volume in Honor of Jacob Marshak*, pages 297–336. cited By 2.

Hurwicz, L. (1973). The design of mechanisms for resource allocation. *The American Economic Review*, 63(2):1–30.

Kagel, J. H., Harstad, R. M., and Levin, D. (1987). Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica*, pages 1275–1304.

Kagel, J. H. and Levin, D. (2016). Auction: A survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *The Handbook of Experimental Economics*, chapter 9, pages 563–637. Princeton University Press, Princeton.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, 59(4):S285–S300.

Kawagoe, T. and Mori, T. (2001). Can the pivotal mechanism induce truth-telling? an experimental study. *Public Choice*, 108(3-4):331–354.

Khalmetski, K. and Sliwka, D. (2017). Disguising lies - image concerns and partial lying in cheating games. *CESifo Working Paper Series No. 6347*.

Köneke, V., Müller-Peters, H., and Fetchenhauer, D. (2015). *Versicherungsbetrug verstehen und verhindern*. Springer.

Kosenok, G. and Severinov, S. (2008). Individually rational, budget-balanced mechanisms and allocation of surplus. *Journal of Economic Theory*, 140(1):126–161.

Köszegi, B. (2014). Behavioral contract theory. *Journal of Economic Literature*, 52(4):1075–1118.

Kucuksenel, S. (2012). Behavioral mechanism design. *Journal of Public Economic Theory*, 14(5):767–789.

Kunreuther, H. (1984). Causes of underinsurance against natural disasters. *Geneva Papers on Risk and Insurance*, pages 206–220.

Lacker, J. M. and Weinberg, J. A. (1989). Optimal contracts under costly state falsification. *Journal of Political Economy*, 97(6):1345–1363.

Laffont, J.-J. and Martimort, D. (2000). Mechanism design with collusion and correlation. *Econometrica*, 68(2):309–342.

Lammers, F. and Schiller, J. (2010). Einflussfaktoren für betrügerisches verhalten von versicherungsnehmern: erste experimentelle befunde. *Zeitschrift für die gesamte Versicherungswissenschaft*, 99(5):649–663.

Lavarreda, S. A., Brown, E. R., and Bolduc, C. D. (2011). Underinsurance in the united states: an interaction of costs to consumers, benefit design, and access to care. *Annual review of public health*, 32:471–482.

Lindsey, C. and Sheather, S. (2010). Variable selection in linear regression. *The Stata Journal*, 10(4):650–669.

Mazar, N., Amir, O., and Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6):633–644.

McAfee, R. P. and Reny, P. J. (1992). Correlated information and mechanism design. *Econometrica*, 60(2):395–421.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

Morgan, J., Steiglitz, K., and Reis, G. (2003). The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy*, 2(1).

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Myerson, R. B. and Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265 – 281.

Nash, J. (1951). Non-cooperative games. *Annals of mathematics*, pages 286–295.

Neeman, Z. (2004). The relevance of private information in mechanism design. *Journal of Economic Theory*, 117(1):55 – 77.

Piketty, T. (1993). Implementation of first-best allocations via generalized tax schedules. *Journal of economic theory*, 61:23–23.

Puelz, R. and Snow, A. (1994). Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market. *Journal of Political Economy*, 102(2):236–257.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302.

Riley, J. G. and Samuelson, W. F. (1981). Optimal auctions. *The American Economic Review*, 71(3):381–392.

Robert, J. (1991). Continuity in auction design. *Journal of Economic Theory*, 55(1):169 – 179.

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

Saito, K. (2013). Social preferences under risk: Equality of opportunity versus equality of outcome. *American Economic Review*, 103(7):3084–3101.

Satterthwaite, M. A. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2):187–217.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55.

Shalvi, S., Gino, F., Barkan, R., and Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*, 24(2):125–130.

Shavell, S. (1979). On moral hazard and insurance. *The Quarterly Journal of Economics*, 93(4):541.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of political economy*, 70(2):111–137.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374.

Stiglitz, J. E. (1977). Monopoly, non-linear pricing and imperfect information: the insurance market. *The Review of Economic Studies*, 44(3):407–430.

Sutter, M. (2009). Deception through telling the truth?! experimental evidence from individuals and teams*. *The Economic Journal*, 119(534):47–60.

Tausch, F., Potters, J., and Riedl, A. (2014). An experimental investigation of risk sharing and adverse selection. *Journal of Risk and Uncertainty*, 48(2):167–186.

Tennyson, S. (1997). Economic institutions and individual ethics: A study of consumer attitudes toward insurance fraud. *Journal of Economic Behavior and Organization*, 32(2):247–265.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37.

von Bieberstein, F. and Schiller, J. (2017). Contract design and insurance fraud: an experimental investigation. *Review of Managerial Science*.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, 16(2):167–207.

Wilson, R. (1987). Game-theoretic analyses of trading processes. In Bewley, T. F., editor, *Advances in Economic Theory: Fifth World Congress*, Econometric Society Monographs, page 3370. Cambridge University Press.

Young, H. P. (2015). The evolution of social norms. *Annual Review of Economics*, 7(1):359–387.

Curriculum Vitae Lars Freund

Persönliche Daten

Geburtsdatum:	25.11.1989
Geburtsort:	Bad Oldesloe, Deutschland
Staatsangehörigkeit:	Deutsch

Universitätsausbildung

seit 10/2015	Promotion im Fach Volkwirtschaftslehre
	Universität zu Köln
10/2013-09/2015	Master of Science in Economics
	Universität Bonn
09/2010-06/2013	Bachelor of Science im Fach Volkwirtschaftslehre
	Universität Mannheim

Akademische Postitionen

seit 10/2015	Research Fellow
	MPI zur Erforschung von Gemeinschaftsgütern, Bonn
04/2014-09/2015	Research Assistent
	MPI zur Erforschung von Gemeinschaftsgütern, Bonn

Ort, Datum Unterschrift

Eidesstattliche Erklärung
nach §8 Abs. 3 der Promotionsordnung vom 17.02.2015

"Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich (zutreffendes unterstreichen) geholfen:

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Ort, Datum Unterschrift