View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Kölner UniversitätsPublikationsServer

Essays in Public Economics

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln
2015

vorgelegt
von
Diplom-Volkswirtin Désirée Rückert

aus

Konstanz

Referent:	Prof. Dr. Felix Bierbrauer
Korreferent:	Prof. Dr. Johannes Münster
Tag der Promotion:	26.10.2015

Acknowledgments

I am deeply indebted to my supervisor Felix Bierbrauer, who has provided me with strong support, guidance and encouragement, throughout all stages of my dissertation. I have constantly profited from his clear and critical feedback and his enthusiasm towards economic research. Chapter 1 and 4 are joint work with him.

I am grateful to Johannes Münster, my second advisor, for spending much effort in numerous discussions and for providing crucial feedback on this thesis. I have benefited from invaluable contributions and advise by Axel Ockenfels, who worked with me on the last chapter of this dissertation. Moreover, I am thankful for great input and enduring impact by Benny Molodvanu and Nora Szech at the Bonn Graduate School of Economics.

It has been a great privilege to co-author the first chapter with Laura Kohlleppel and the fourth chapter with Andreas Pollak. Our interaction were a constant source of learning for me. I thank them both for their absolute commitment to our common projects.

All chapter of this thesis have benefited from various discussions and comments. In this respect I like to thank Martin Barbie, Pierre Boyer, Dirk Engelmann, Alina Gizatulina, Jacob Goeree, Hans-Peter Grüner, Christian Hellwig, Stephen Morris, John Weymark and anonymous referees. Emanuel Hansen and Dominik Sachs provided crucial input to this thesis by reading, listening to and commenting on my research projects.

I would also like to thank my fellow graduate students and friends in Cologne and Bonn for their continuous support. Their encouragement and understanding were invaluable contribution to the success of this thesis. Sinem Hacioglu, Benjamin Enke, Ralph Lütticke, Nadin Heß and Christian Geppert were a particularly important source of energy. I am also greatful to Kerstin Burghaus, Christopher Busch, Moritz Drexl, Rafael Flore, Max Groneck, Benjamin Guin, Thorsten Kemper, Andreas Kleiner, Klaudia Michalek-Kursawe, Lien Pham-Dao, Öznur Saglam, Matthias Schön, Frederik Schwerter, Frederik Thenée and Christoph Wigger for providing me with their company and giving me such a nice time during my PhD.

I would like to express my deepest gratitude towards my parents, Martin and Silvia, for their unconditional support throughout my whole life and everything they did for my personal and professional development. Finally, and most importantly, I thank Nicolas for his love and for unselfishly standing by my side in all the time.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of four independent chapters which are linked in several aspects: All chapters contribute to the theory of public economics. From a theoretical point of view, all chapters are based on the assumption that agents are privately informed about their preferences, and all chapters use mechanism design theory. Yet their applications vary and cover topics such as public good provision, externality regulation and income taxation. The first three chapters form an entity as they use the independent private values model. Chapter 4 uses robust mechanism design.

Chapter 1 studies the independent private values model in mechanism design, applied to the problem of bilateral trade and public good provision. It provides conditions under which a model with a large but discrete number of types behaves qualitatively in the same way as a model with a continuum of types. Chapters 2 and 3 deal with the problem of externality regulation. I consider firms that can reduce externalities, which is beneficial to consumers. Firms have private information about their costs, and consumers have private information about their preferences. Chapter 2 investigates optimal price instruments (e.g. taxes) and quantity instruments (e.g. tradable permits). These two instruments are frequently used to regulate externalities such as CO_2-emissions, acid rain and water pollution. Both instruments are contrasted with the optimal unconstrained mechanism to regulate externalities. Chapter 3 addresses the question how externalities should be regulated when distributional concerns and efficiency are considered. If stronger weight is put on consumers in the regulator's welfare function, lower emission reduction takes place than when the regulator is interested in jointly maximizing consumer surplus and firm profits. Chapter 4 varies in that it allows preferences to be different from selfish. It is a contribution to the theory of robust mechanism design, taking into account findings of experimental research. More precisely, it describes how mechanisms can be designed that are not only robust with respect to variations in beliefs but as well to deviations from standard preferences.

Chapter 1 The first chapter is based on a joint project with Felix Bierbrauer and Laura Kohlleppel. It studies the independent private values model, a workhorse model of mechanism design. Consumers are privately informed about their preferences, and firms are privately informed about their costs. This framework covers many applications such as bilateral trade, auctions, public good provision or partnership dissolution. Typically, the independent private values model is based on the assumption that consumers and firms have a continuum of types. It works with the assumption of an atomless distribution. We introduce an alternative specification and assume that agents have a discrete number of types.

For this model specification, we derive necessary and sufficient conditions for the possibility to implement a social choice function. For our characterization of implementable outcomes, we introduce a measure of how difficult it is to implement a given social choice function. We use this measure to provide comparative static results. We ask, for instance, whether an increased number of types or an increased number of agents make it more difficult to implement efficient outcomes. In particular, we discuss the discrete type analogues to the impossibility result by Myerson and Satterthwaite (1983) for the bilateral trade problem, and by Mailath and Postlewaite (1990) for the public good provision problem. We find that the Mailath and Postlewaite result extends to any model with a discrete set of types. By contrast, for the Myerson and Satterthwaite result, we find parameter constellation such that efficient bilateral trade is possible if the number of types for the buyer and the seller is small. A final contribution of this chapter is that it provides conditions under which a model with a large but discrete number of types behaves approximately in the same way as the model with a continuum of types.

Chapter 2 In the second chapter, I apply the independent private values model developed in Chapter 1 to analyze externality regulation when the regulatory agency has different instruments at hand to achieve socially optimal results. More precisely, I introduce a problem of emission reduction into the independent private values model: There are firms that can reduce emissions, and they are privately informed about their costs to achieve the reduction. The emission reduction benefits consumers that have private information about how much they value emission reduction. Following the literature on externality regulation under uncertainty, I assume that the regulating agency has a price instrument (e.g. a tax) or a quantity instrument (e.g. tradable permits) at hand to regulate emissions in order to maximize social surplus.[1]

The characterization of optimal price and quantity instruments for externality regulation is treated as a problem of mechanism design. For this purpose, I assume that the mechanism designer introduces price and quantity mechanisms before private information is revealed. I compare these two mechanism with an optimal unconstrained mechanism, i.e. a mechanism that avoids any a priori assumption on the set of admissible policies. In this context, I show that the unconstrained mechanism leads to ex post surplus maximization, where overall surplus consists of consumer surplus and firm surplus. The price and quantity mechanisms, by contrast, fail to achieve ex post efficiency. The Coase theorem would hence suggest that an optimally designed mechanism is able to improve upon both price and quantity instruments (Coase, 1960). I show that this

[1]This line of research starts with the seminal paper of Weitzman (1974).

logic does not apply here. There is always one type that is better off under the already installed mechanism (i.e., price or quantity mechanisms) than under the optimal unconstrained mechanism.

Chapter 3 The third chapter is based on the same environment as Chapter 2. There are firms that can reduce emissions, which is beneficial to consumers. All agents privately observe the realization of their characteristics. The regulating agency can use price and quantity mechanisms to reduce emissions. The difference to Chapter 2 lies in the surplus function of the regulating agency. In Chapter 3, I assume that the regulator is interested in maximizing consumer surplus as opposed to total surplus that consists of consumer surplus and firms' profits. This allows me to investigate how the optimal level of emission reduction is affected by distributional considerations. When the regulating agency maximizes consumer surplus, less emission reduction takes place than when the regulating agency maximizes total surplus. This is surprising because consumers, who are harmed by emissions, prefer less emission reduction than the total surplus maximizing amount.

This observation is independent of the regulator's choice of mechanism. The optimal unconstrained mechanism, as well as optimal price and quantity mechanisms lead to less emission reduction under consumer surplus maximization. However, the mechanisms differ with respect to efficiency and surplus distribution under the two objectives. Comparative static properties of the solution to optimal externality regulation under the three mechanisms are provided. The parameters for which comparative statics are consumers' preference parameter and firms' cost parameter.

Chapter 4 The fourth chapter of the thesis is a modified version of a joint paper with Felix Bierbrauer, Axel Ockenfels and Andreas Pollak (Bierbrauer et al., 2015). It is motivated by observations from behavioral economics. In laboratory experiments, agents deviate from selfish behavior. These deviations can be explained by social preferences as inequality aversion, altruism or intentionality. In standard mechanism design theory, this is not taken into account. Thus, mechanisms that are robust in the sense that they do not rely on a common prior distribution of material payoffs, might not be robust to variations in preferences, in particular, that individuals are motivated by social preferences.

In this chapter, it is shown how social preferences can be taken into account in robust mechanism design. Two classics in mechanism design are studied: Trade between two parties with private information on their valuations, and redistribution among agents with private information on their productive abilities – to show that some, but not all, standard mechanism design solutions fail with social preferences. We characterize optimal mechanisms for the bilateral trade problem and the problem of redistributive income taxation under selfish preferences and provide laboratory evidence that a non-negligible share of individuals deviates from the behavior that would be predicted on basis of selfish preferences. We show that this can be explained by models of social preferences. We introduce the notion of social-preference-robust mechanisms, which allows to control behavior not only for selfish but also for social preferences of different nature and intensity, and characterize the optimal mechanism in this class. We present laboratory evidence that these mechanisms successfully control behavior.

Finally, we compare the performance of the optimal mechanism for selfish agents and the social-preference robust mechanism with the help of laboratory experiments. We find that behavior can indeed be better controlled with social-preference-robust mechanisms. However, the ability to control behavior is not the same as the ability to reach a given objective. In our analysis of the bilateral trade problem, a mechanism is designed with the objective of profit-maximization. In our experimental data, profits are higher with the mechanism that is designed for individuals with selfish preferences. Contrary, for the analysis of welfare-maximizing income taxation the social-preference-robust mechanism leads to higher welfare.

1
On the independent private values model – A unified approach

1.1 Introduction

The independent private values model is an important workhorse model for the theory of mechanism design. In this model, economic agents are privately informed about their characteristics, typically preferences or costs, and, moreover, the characteristics of different agents are modeled as the realizations of independent random variables. In addition, an individual's payoff does not depend on the types of other individuals. This framework has been applied to study a wide range of allocation problems. These include the allocation of indivisible private goods (auctions), the provision of pure or excludable public goods, the regulation of externalities, the problem of partnership dissolution, or redistributive income taxation.

The seminal papers in this literature are based on the assumption that, for each agent, there is a continuum of possible types and that the corresponding probability distribution has no mass points and a monotone hazard rate. Moreover, the typical approach is to use the envelope theorem for a characterization of incentive compatible social choice functions. In this paper, we develop an alternative characterization of implementable social choice functions that is based on the assumption that the set of possible types is discrete. More specifically, our analysis proceeds as follows:

We first provide necessary and sufficient conditions for the implementability of a social choice function. For our characterization, we introduce the notion of a *minimal* *subsidy*. It is defined as the difference between the maximal payment that one can extract from individuals in the presence of incentive and participation constraints, and the payment that would be required in order to ensure budget balance. That is to say, the minimal subsidy is the amount of money an external party would have to provide so as to make a given social choice function compatible with the requirements of incentive compatibility, voluntary participation, and budget balance. However, we do not assume that such an external party is actually available. Consequently, a social choice function can be implemented if and only if the minimal subsidy is negative. We then apply our characterization to clarify conditions under which the famous impossibility results by Myerson and Satterthwaite (1983) and Mailath and Postlewaite (1990) extend to a model with a discrete set of types. Second, we provide a comparative statics analysis of how a change in exogenous parameters – such as the number of individuals, or the number of possible types per individual – affect the minimal subsidy. This allows us, for instance, to check whether a change in the economic environment makes it more or less difficult to implement an efficient provision rule for public goods. A final contribution of our paper is to spell out the conditions under which a model with a large but discrete number of types behaves approximately in the same way as a model with a continuum of types.

These results are derived in a model in which many consumers, who have private information about their preferences, benefit from the provision of a private or public good. Their payoffs are quasi-linear in the transfers they need to pay for the good. Further, many firms, which have private information about their costs, profit from the production of goods. Firm profits are quasi-linear in the revenues they receive for producing the good. Consumers' consumption is bounded by the total output that is made available by the firms. To derive the minimal subsidy, we proceed as follows: The social choice function can be divided into a transfer and consumption rule for consumers and a revenue and production rule for firms. First, we hold the consumption rule for consumers fixed and derive the maximal transfers that consumers are able to make if incentive compatibility constraints and participation constraints need to be respected simultaneously. Similarly, we hold the production rule fixed for firms and derive the minimal revenue that firms are willing to accept if again incentive compatibility constraints and participation constraints need to be respected. The differences between the maximal consumer transfers and the minimal firm revenues is the minimal subsidy. If the minimal subsidy is positive, i.e., the mechanism runs a deficit, then the specified consumption and production rules are not implementable. Contrary, if the minimal subsidy is negative, the implementation of the social choice function is possible. For the characterization of maximal consumer transfers and minimal firm revenues we use techniques developed in the non-linear pricing literature (e.g. Bolton and Dewatripont, 2005).

Our analysis proceeds as follows: We first derive necessary and sufficient conditions for the implementation of a social choice function. For the characterization of the first condition, we consider the problem of maximizing consumers' transfers and the problem of minimizing firms' revenues, taking only a subset of incentive compatibility and participation constraints into consideration. Specifically, we consider the participation constraints of the consumer with the lowest valuation for consumption and the incentive constraints that prevent consumers to communicate lower preferences. Similarly, we take into consideration the participation constraint of the firm with the highest costs of production and the incentive constraints that prevent firms from exaggerating their costs. The expression that arises from this *relaxed* problem, where only a subset of constraints is considered, provides a lower bound on the minimal subsidy. Thus, a necessary condition for the implementation is that the minimal subsidy of this relaxed problem is negative. Second, we derive a sufficient condition, which assures that the lower bound of the minimal subsidy can be reached. This condition requires that the consumption rule and the production rule are monotone, so that consumers with higher willingness to pay for the good consume more than consumers with a lower willingness to pay; and similarly, firms with lower costs produce more output than firms with higher costs.

These conditions have the following implications: First-best consumption and provision rules are monotone. Therefore, first-best implementation is possible if and only if the minimal subsidy is negative. When the first-best provision rule is not implementable, monotonicity of the consumption and provision rules can be achieved when the distribution of agents' types satisfies a monotone hazard rate assumption. Hence, consumption and production plans that maximize a social surplus function subject to the constraint that the minimal subsidy is negative, are monotone and therefore implementable. To derive the necessary condition, the monotonicity of hazard rates does not play a role. We present a version of the impossibility results of Mailath and Postlewaite (1990), when consumers have a discrete number of types. Our specification uses only the necessary condition to derive this result. We do not require the assumption of a monotone hazard rate that was imposed by Mailath and Postlewaite (1990) in order to attain the impossibility result. Hence, our result holds under less restrictive assumptions.

We provide comparative static results that show how the minimal subsidy varies with the number of types and the number of agents. In particular, we can compare the comparative static properties of the minimal subsidy in a private good setting with a public good setting. A change in the number of agents affects the minimal subsidy in both settings differently. In a public good setting, an increase in the number of consumers leads to a positive minimal subsidy, so that it is impossible to efficiently provide the public good. Contrary, when a private good setting is considered, an increase in the number of buyers and sellers leads to a negative minimal subsidy. An increase in the number of types, on the other hand, increases the minimal subsidy, so that in the private good setting, as well as the public good setting, impossibility results occur when the number of types grows large.

In order to understand how parameter changes affect the minimal subsidy, we decompose the effect of a change in parameters in the surplus measure and the measure for information rents separately. We show that when each agent has a binary type set, then parameters can be found such that efficient bilateral trade is possible. Further, if only two consumers are considered, parameters can be found such that the public good can be provided efficiently. We show that the ability to reach possibility results hinges on the observation that parameters need to be chosen in such a way that the surplus measure is bigger than the information rents that need to be guaranteed. This raises the question how (i) the agent's type parameters, (ii) the probability weights on types (iii) the number of types and (iv) the number of agents influence the minimal subsidy. In particular, we show that the possibility results that are derived with a binary type set 'approach' the impossibility results of Myerson and Satterthwaite (1983) and Mailath and Postlewaite (1990) if the number of types increases and if the new types are introduced in such a way that the finite type set lies *dense* in the infinite type set, i.e., every point of the infinite type set can be approximated by a point of the finite subset of types. We demonstrate that the minimal subsidy in the discrete setting converges to the minimal subsidy in the continuous setting if the environments are aligned.

Based on these observations, we study general convergence results. We specify what we mean by one environment approaches another environment, so that results in a continuous setting and a discrete setting coincide. Therefore we define an environment that allows us to compare and relate different economies, e.g. the discrete and the continuous bilateral trade economy. Each of the economies is characterized by four decisive factors for implementability: the number of agents, the number of types, the probability distribution and the parameter constellation. Formally, we can approximate 'similar' economies by adjusting the single components; i.e., as we increased the number of types and adjusted the probability distribution, we transfer the discrete bilateral trade setting into the setting of Myerson and Satterthwaite (1983). If the components are adjusted 'appropriately', we say that one economy will converge to the other economy. To relate and analyze implementability results of different settings, we calculate the minimal subsidy and study what drives the possibility to attain efficient implementation in each economy. We give general insights on how different applications of the independent private values can be linked.

The reminder is organized as follows. The next section contains a more detailed discussion of the literature. Section 1.3 provides counterparts to Myerson and Satterthwaite (1983) and Mailath and Postlewaite (1990) impossibility results. Section 1.4 introduces the model. To motivate our more general analysis in subsequent sections, Section 1.5 presents necessary and sufficient conditions for the implementation of social choice functions and characterizes first-best and second-best provision rules. Section 1.6 discusses comparative statics properties of the bilateral trade problem and the public good provision problem. Further, this section studies the impact of increasing the number of agents on the possibility result. Section 1.7 then shows how the discrete type setting converges to the continuous type setting and analyzes the convergence in our examples.

The last section contains concluding remarks. Preliminary proofs are in part 1.A of the Appendix. Part 1.C introduces further applications.

1.2 Related literature

The independent private values model has been applied to study a wide variety of allocation problems, from the allocation of indivisible private goods (auctions), to the bilateral trade problem, the provision of pure or excludable public goods, the regulation of externalities, the regulation of a monopolist, or the problem of partnership dissolution.

In auction theory, the independent private values model is central. The seminal paper that introduced the second-price auction and the revenue equivalence theorem is Vickrey (1961). Optimal auctions for risk neutral bidders with independent types are derived in Myerson (1981), Riley and W. (1981), Harris and Raviv (1981) (see McAfee and McMillan, 1987, for further references). Che and Gale (2006) show that the revenue equivalence theorem does not need to apply when the number of buyer types is finite.

Further, for the example of bilateral trade, Myerson and Satterthwaite (1983) have shown that if the buyer's preferences and the seller's costs are private information and voluntary participation needs to be assured, efficient trade is not possible. They introduce the notion of the minimal subsidy and thereby provide a measure of how severe the impossibility result is. We will show that this impossibility result does translate into a model with many finite types. For few finite types, however, parameters can be found such that efficient bilateral trade is possible; i.e., the minimal subsidy is negative. A special case of our setup is the paper of Matsuo (1989), who provides conditions under which efficiency in the bilateral trade example can be reached for discrete distributions with two types.

The possibility to achieve efficient public good provision, as a Bayes-Nash equilibrium in an independent private values model, has first been established by D'Aspremont and Gerard-Varet (1979) and Arrow (1979). This literature has not taken voluntary participation into account. Güth and Hellwig (1986), Rob (1989) and Mailath and Postlewaite (1990) have shown that if preferences for public goods are private information, so that incentive compatibility constraints need to be considered and if at the same time voluntary participation need to be guaranteed, then first-best efficient public good provision cannot be achieved. We will show that this impossibility result does not rely on the assumption that preferences for public goods are continuously distributed by showing that the impossibility extends to a setup with an arbitrary discrete type set. The provision of a non-rival, but excludable good is studied in Güth and Hellwig (1986), Hellwig (2003), Schmitz (1997) and Norman (2004).

The independent private values model has also been applied to study the dissolution of partnerships, see Cramton, Gibbons, and Klemperer (1987). They look at situations where each of several agents possesses a fraction of a good and assume a continuous symmetric distribution of agents' valuations. They show that an efficient reassignment of shares is possible if initial shares are sufficiently equal distributed. However, when a single agent possesses all shares of the partnership, then the same arguments as in Myerson and Satterthwaite (1983) apply and an efficient dissolution is impossible. Hence, whether the partnership can be dissolved efficiently relies on the initial shares of the partnership. As a corollary of our analysis of Myerson and Satterthwaite (1983) with discrete types, we show that if the number of types is finite, then parameters can be found such that the partnership can always be dissolved efficiently, even when shares are unevenly distributed.

Hellwig (2007) provides separate characterizations of optimal income taxes for a model with a discrete set of types and for a model with a continuum of types. He argues that for all steps in the proof for the continuous type set there exists an analogous step for the discrete type set. The strategy of our paper is different in that we analyze the implementation of social choice functions for an arbitrary number of discrete types. We investigate the implication of this modeling choice by approximating the continuous type set.

Kos and Messner (2013) provide a general characterization of implementable allocation rules. They describe bounds on the set of transfers that implement an allocation rule. They do refrain from any assumption on the agent's type set and utility function. The work of Kos and Messner (2013) is related to this paper in that it makes use of minimal subsidies to evaluate whether implementation of social choice functions is possible. Opposed to them, we make more specific assumptions that allow us to elaborate more clearly on necessary and sufficient conditions for implementation. Further, it enables us to do comparative statics analysis.

The specification of the independent private values model with a finite number of types is well suited for directly testing mechanisms in the laboratory. Bierbrauer et al. (2015) use this specification to test whether mechanism that are robust to agents' probabilistic beliefs (see Bergemann and Morris, 2005) fail when agents have social preferences.

1.3 Motivating examples

This section contains motivating examples, which illustrate the difficulty of extending the impossibility results by Myerson and Satterthwaite (1983) and Mailath and Postlewaite (1990) to models with a discrete set of types. Throughout, we will use these two examples to illustrate conceptual issues that arise.

1.3.1 The bilateral trade problem

In the private good setting, there is one buyer and one seller. The seller produces $y \in [0,1]$ units of a good. The buyer can purchase $q \in [0,1]$ units of the good. The buyer's utility is given by $u(\theta, q, t) = \theta q - t$, so that θ is the buyer's valuation for the good and t is the transfer the buyer has to pay for the good. The seller's profit is given by $\pi(\delta, y, r) = r - \delta y$, so that δ is the cost of producing the good and r is the revenue the seller receives for providing the good. The quantity that is consumed by the buyer is equal to the quantity produced by the seller, so that for all $(\theta, \delta) \in \Theta \times \Delta$, $q(\theta, \delta) = y(\theta, \delta) \in [0, 1]$. Further, it is assumed that trade is voluntary and, in the absence of trade, both parties realize a utility, respectively a profit of 0. We define for the buyer a function $Q : \Theta \mapsto [0, 1]$, where $Q(\theta^k) = \mathbb{E}_{(\delta)}[q(\theta^k, \delta)|\theta^l]$. This gives the conditional expectation over the probability that the buyer gets the good, in case that he announces type θ^k but having a true type θ^l. The conditional expected value of the transfers $T(\theta^k)$, and the conditional expected values of revenues $R(\delta^k)$ and produced quantity $Y(\delta^k)$ for the seller are defined analogously. The seminal analysis of the bilateral trade problem by Myerson and Satterthwaite (1983) has focused on the question whether there exists a Pareto efficient or surplus-maximizing social choice function that is incentive-compatible for the buyer,

$$\theta^l Q(\theta^l) - T(\theta^l) \geq \theta^l Q(\theta^k) - T(\theta^k) \ , \qquad \forall \, \theta^l, \theta^k \in \Theta \ , \tag{1.1}$$

incentive-compatible for the seller,

$$R(\delta^l) - \delta^l Y(\delta^l) \geq R(\delta^k) - \delta^l Y(\delta^k) \ , \qquad \forall \, \delta^l, \delta^k \in \Delta \ , \tag{1.2}$$

and compatible with the budget requirement,

$$\mathbb{E}_{(\theta,\delta)} \left[t(\theta, \delta) \right] \geq \mathbb{E}_{(\theta,\delta)} \left[r(\theta, \delta) \right] . \tag{1.3}$$

Surplus-maximization requires that the function $q : \Theta \times \Delta \to [0,1]$ is chosen so as to maximize

$$\mathbb{E}_{(\theta,\delta)} \left[(\theta - \delta) q(\theta, \delta)) \right] \ .$$

Hence, surplus-maximization requires that

$$q(\theta, \delta) = \begin{cases} 0, & \text{if} \quad \theta < \delta \ , \\ 1, & \text{if} \quad \theta > \delta \ . \end{cases}$$

Myerson and Satterthwaite (1983) analyzed the bilateral trade problem under the assumption of an atomless distribution functions with a monotone hazard rate[1], and es-

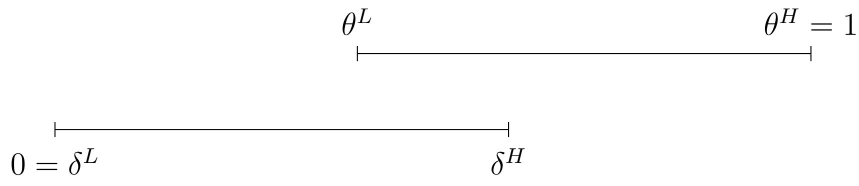
[1]The hazard rate for the buyer is defined as $\frac{1-F(\theta)}{f(\theta)}$, where the cumulative distribution function of the

tablished the following impossibility result.

Proposition 1.1. ***Myerson and Satterthwaite (1983):*** *If the buyer's valuation for the good is independently drawn from the intervals $[\theta^L, \theta^H]$ and the seller's costs for the good are drawn from the interval $[\delta^L, \delta^H]$ with strictly positive densities, such that the interiors of the intervals $[\theta^L, \theta^H]$ and $[\delta^L, \delta^H]$ are not disjunct, then there is no Bayesian incentive compatible social choice function that is ex post efficient and gives every buyer type and every seller type non-negative expected gains from trade.*

We change the assumption of atomless type distributions and show: When the buyer's and the seller's type set is discrete, then efficient trade is possible for some parameter constellations.

Figure 1.1: Binary type set



Assume that each agent's type occurs with equal probability. For certain parameter constellations, e.g.

$$\underline{\theta}_s = 0 \ , \qquad \underline{\theta}_b = \frac{1}{8} \ , \qquad \bar{\theta}_s = \frac{7}{8} \qquad \text{and} \qquad \bar{\theta}_b = 1 \ ,$$

the Myerson and Satterthwaite (1983) impossibility result is obtained. Consider the following *relaxed problem*: The mechanism designer is interested in maximizing expected surplus, subject to the incentive compatibility constraints in (1.1) and (1.2) and subject to the constraints that gains from trade have to be non-negative. The following Table gives a solution to this relaxed problem.

Table 1.1: Positive minimal subsidy

(q, r, t)	$\underline{\theta}_s$	$\bar{\theta}_s$
$\underline{\theta}_b$	$\left(1, \frac{3}{8}, \frac{1}{8}\right)$	$(0, 0, 0)$
$\bar{\theta}_b$	$\left(1, \frac{4}{8}, \frac{4}{8}\right)$	$\left(1, \frac{7}{8}, \frac{5}{8}\right)$

random variable is denoted by F and f is the density; for the seller, the hazard rate is defined as $\frac{P(\delta)}{p(\delta)}$, respectively.

The maximal expected transfer that the buyer is willing to make is $\mathbb{E}_{(\theta,\delta)}\left[t(\theta,\delta)\right] = \frac{5}{16}$. The minimal expected revenue the seller is willing to accept is $\mathbb{E}_{(\theta,\delta)}\left[r(\theta,\delta)\right] = \frac{7}{16}$. Hence, the solution to the relaxed problem violates the budget constraint in (1.3). The minimal subsidy that is necessary for efficient bilateral trade is $\mathbb{E}_{(\theta,\delta)}\left[r(\theta,\delta)\right] - \mathbb{E}_{(\theta,\delta)}\left[t(\theta,\delta)\right] = \frac{1}{8}$.

Contrary, for the following parameter constellation

$$\underline{\theta}_s = 0 \ , \qquad \underline{\theta}_b = \frac{1}{3} \ , \qquad \bar{\theta}_s = \frac{2}{3} \qquad \text{and} \qquad \bar{\theta}_b = 1 \ ,$$

the social choice function, which specifies (q, t, r) for all possible type combinations, in Table 1.2 leads to efficient bilateral trade and satisfies the conditions in (1.1), (1.2), (1.3) and assures non-negative payoffs.[2] Whenever the buyer's marginal valuation for the good is higher than the seller's marginal costs, the good is exchanged. The maximal expected transfer that the buyer is willing to make is $\mathbb{E}_{(\theta,\delta)}\left[t(\theta,\delta)\right] = \frac{1}{2}$. The minimal expected revenue the seller is willing to accept is $\mathbb{E}_{(\theta,\delta)}\left[r(\theta,\delta)\right] = \frac{1}{3}$. Hence, the budget constraint in (1.3) is satisfied. The minimal subsidy that is necessary for efficient bilateral trade is $\mathbb{E}_{(\theta,\delta)}\left[r(\theta,\delta)\right] - \mathbb{E}_{(\theta,\delta)}\left[t(\theta,\delta)\right] = -\frac{1}{6}$.

Table 1.2: Negative minimal subsidy

(q, r, t)	$\underline{\theta}_s$	$\bar{\theta}_s$
$\underline{\theta}_b$	$\left(1, \frac{1}{3}, \frac{2}{3}\right)$	$(0, 0, 0)$
$\bar{\theta}_b$	$\left(1, \frac{1}{3}, \frac{2}{3}\right)$	$\left(1, \frac{2}{3}, \frac{2}{3}\right)$

1.3.2 Public good provision

An indivisible public good is either provided or not. There are $I = \{1, \ldots, n\}$ consumers and one producer. The utility function is taken to be linear so that $u(\theta_i, q, t) = \theta_i \, q - t$, where $q = 1$ if the public good is provided, and $q = 0$, otherwise. The producer's cost function is taken to be publicly known. If the public good is produced, the costs are equal to nc, where c is the per capita cost of public-goods provision. Since the cost function is known, the producer's incentive compatibility constraints are irrelevant, and a state of the economy is exclusively defined by the vector of preference parameters θ. When the public good is not provided, all consumers realize a utility of zero.

The analysis of public good provision by Mailath and Postlewaite (1990) has focused on the question whether there exists a Pareto efficient social choice function that is

[2]See Observation 1.1 below, for a proof.

incentive compatible, so that the incentive constraints in (1.1) are satisfied for all i, and that satisfies the resource requirement

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n} t_i(\theta)\right] \geq nc\,\mathbb{E}_{(\theta)}[q(\theta)]\ , \tag{1.4}$$

where n denotes the number of consumers.

Surplus maximization requires that the function $q : \Theta^n \mapsto [0,1]$ is chosen so as to maximize

$$\mathbb{E}_{(\theta)}\left[\left(\sum_{i=1}^{n} \theta_i - nc\right) q(\theta)\right]\ .$$

Hence, surplus maximization requires that

$$q(\theta) = \begin{cases} 0, & \text{if } \frac{1}{n}\sum_{j=1}^{n}\theta_i < c\ , \\ 1, & \text{if } \frac{1}{n}\sum_{j=1}^{n}\theta_i > c\ . \end{cases}$$

If the average valuation of a consumer exceeds the per capita costs, $\mathbb{E}_{(\theta)}[\theta_i] > c$, then the public good should be provided. Mailath and Postlewaite (1990) analyze the public good provision under the assumption of atomless distribution functions with a monotone hazard rate. They show that if the number of consumers grows without limit, public good provision is zero under any social choice function that is incentive compatible and respects participation constraints.

Proposition 1.2. ***Mailath and Postlewaite (1990):*** *If the consumers' valuation for the public good are independently drawn from the intervals $[\theta^L, \theta^H]$ with strictly positive densities and the per capita costs are such that $\theta^L < c < \theta^H$, then $lim_{n\to\infty} prob\,(q(\theta)^n > 0) = 0$, for any mechanism in the sequence of mechanism satisfying incentive compatibility constraints, voluntary participation and expected budget balance.*

With many consumers, if the average valuation is higher than the marginal per capita costs, then the efficient amount of public good provision will be almost surely equal to 1, yet the amount that is going to be implemented will be almost surely equal to 0, under any mechanism that respects consumers' voluntary participation.

Consider now the case where consumers have a binary type set and assume that each consumer's type occurs with equal probability. For all parameter constellations $\theta_i^L < c < \theta_i^H$, there is no social choice function, which maximizes surplus and satisfies incentive compatibility constraints, ensures non-negative utility for all consumers and fulfills the budget requirement (see Proposition 1.5 below). With private information on public good preferences, consumer i's transfers have to be chosen such that the incentive compatibility constraints in (1.1) are satisfied. However, when the number of consumers grows large, a consumer's impact on the public good provision becomes insignificant. If no consumer has an impact on the provision, then incentive compatibility constraints imply that the transfers have to be similar. Thus, the maximal transfers per capita is θ_i^L, which is smaller than the per capita costs of public good provision.

1.3.3 Comparison of private and public good.

Assume that there are two consumers and that each consumer type occurs with equal probability. For the following parameters, $\underline{\theta}_b = 1,\ c = 3$ and $\bar{\theta}_b = 10$, there is a social choice function, which maximizes surplus, satisfies incentive compatibility constraints in (1.1), assure voluntary participation and fulfills the budget requirement in (1.4). By contrast, for $\underline{\theta}_b = 1,\ c = 3$ and $\bar{\theta}_b = 6$, there is no social choice function, which maximizes surplus and fulfills all constraints.[3] It depends on the parameters whether we have an impossibility result or not. With many individuals, the Mailath and Postlewaite (1990) result extends to a model with a discrete type set. As has been shown by Gresik and Satterhwaite (1989), the Myerson and Satterthwaite (1983) result does not extend to a model with a large number of buyers and sellers. This raises the following more general questions: What impact does the number of agents have on the impossibility results, and what impact does the assumption on the type set have? To address these questions we will develop a general framework in the subsequent section.

1.4 The model

Consumers. There is a finite set of consumers, $I = \{1, ..., n\}$. The preferences of consumer i are represented by the utility function

$$u_i(\theta_i, q_i, t_i) = v(\theta_i, q_i) - t_i \ ,$$

where q_i denotes i's consumption of a public or private good and the function v gives the utility of consumption. It depends on a preference parameter θ_i that belongs to a finite ordered set of possible preference parameters $\Theta_i = \{\theta_i^0, \theta_i^1, ..., \theta_i^s\}$, with $\theta_i^0 < \theta_i^1$, etc. for every $i \in I$. The monetary payment of consumer i is denoted by t_i.

The function v is assumed to have the following properties. Zero consumption gives zero utility: for all $\theta_i \in \Theta$, $v(\theta_i, 0) = 0$. The lowest type does not benefit from consumption: for all q_i, $v(\theta_i^0, q_i) = 0, \forall\ i \in I$. For all other types, the marginal benefit from increased consumption is positive and decreasing, so that for all $\theta_i > \theta_i^0$ and all q_i, $v_2(\theta_i, q_i) > 0$[4] and $v_{22}(\theta_i, q_i) \leq 0$. The marginal benefit of consumption is increasing in the individual's type, so that $\theta_i' \geq \theta_i$ implies that $v_2(\theta_i', q_i) \geq v_2(\theta_i, q_i)$.

[3] See Observation 1.4 below, for a proof.

[4] The index 2 denotes the partial derivative with respect to the second argument; $v_2(\theta_i, q_i) = \frac{\partial v(\theta_i, q_i)}{\partial q_i}$.

The consumer privately observes θ_i. From the perspective of all other agents it is a random variable with support Θ_i and probability distribution $f_i = (f_i^0, ..., f_i^s)$. The random variables $(\theta_i)_{i\in I}$ are independently and identically distributed *(i.i.d.)*. We write $\theta = (\theta_1, ..., \theta_n)$ for a vector of all consumers' taste parameters and θ_{-i} for a vector that lists all taste parameters except θ_i.

For later reference we introduce the following notation: We denote by $f_i(\theta_i)$ a random variable that takes the value f_i^l if θ_i takes the value θ_i^l and by $F_i(\theta_i)$ a random variable that takes the value $\sum_{k=0}^{l} f_i^k$, if θ_i takes the value θ_i^l. Also, we denote by θ_i^+ a random variable that takes the value θ_i^{l+1} if θ_i takes the value θ_i^l, for $l \in \{0, \dots, s-1\}$. If $\theta_i = \theta_i^s$, then the value of θ_i^+ is some arbitrary number.

Producers. There is a set of producers, $J = \{1, ..., m\}$. Each producer contributes to the supply of a public or private good. The contribution of producer j is denoted by y_j and comes with production costs $k(\delta_j, y_j)$, where δ_j is a cost characteristic of firm j that belongs to the finite ordered set $\Delta_j = \{\delta_j^1, ..., \delta_j^r\}$ of possible technology parameters. We assume that $\delta_j^1 < \delta_j^2$ etc. $\forall\, j \in J$. The profit of producer j is given by

$$\pi_j(\delta_j, r_j, y_j) = r_j - k(\delta_j, y_j)\ ,$$

where r_j is producer j's revenue, or, equivalently, a monetary payment to producer j.

The function k is assumed to have the following properties. Zero production is costless: for all $\delta_j \in \Delta_j$, $k(\delta_j, 0) = 0$. The marginal costs from increased production is positive and increasing, so that for all δ_j and all y_j, $k_2(\delta_j, y_j) > 0$ and $k_{22}(\delta_j, y_j) \geq 0$. The marginal cost of production is increasing in the firm's type, so that $\delta_j' \geq \delta_j$ implies that $k_2(\delta_j', y_j) \geq k_2(\delta_j, y_j)$.

The technology parameter δ_j is privately observed by producer j. From the perspective of all other agents, it is a random variable with support Δ_j and probability distribution $p_j = (p_j^1, ..., p_j^r)$. The random variables $(\delta_j)_{j\in J}$ are *i.i.d.* We write $\delta = (\delta_1, ..., \delta_m)$ for a vector of technology parameters and δ_{-j} for a vector that lists all technology parameters except δ_j.

We denote by $p_j(\delta_j)$ a random variable that takes the value p_j^l if δ_j takes the value δ_j^l and by $P_j(\delta_j)$ a random variable that takes the value $\sum_{k=1}^{l-1} p_j^k$ if δ_j takes the value δ_j^l, for $l > 1$, and $P_j(\delta_j)$ takes the value 0 if $\delta_j = \delta_j^1$. We denote by δ_j^- a random variable that takes the value δ_j^{l-1} if δ_j takes the value δ_j^l, for $l \in \{2, \dots, r\}$. If $\delta_j = \delta_j^1$, then the value of δ_j^- is some arbitrary number.

The consumers' preference parameters and the firms' cost parameters are taken to to be independent random variables. We will also refer to a vector (θ, δ) that lists all taste and cost parameters as a state of the economy. The set of all states is given by $(\Theta_i^n)_{i\in I} \times (\Delta_j^m)_{j\in J}$.

Social choice functions/ Direct Mechanisms. A social choice function or direct mechanism consists of a consumption and a payment rule for each consumer i and a production and revenue rule for each producer j. The consumption rule is a function $q_i : \Theta^n \times \Delta^m \mapsto \mathbb{R}_+$, that assigns to each state of the economy a consumption level for consumer i. Analogously, $t_i : \Theta^n \times \Delta^m \mapsto \mathbb{R}$ specifies i's payment as a function of the state of the economy. The production and revenue rule for producer j are, respectively, given by $y_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}_+$ and $r_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}$. We also write $q = (q_i)_{i \in I}$ for the collection of all consumption rules, $y = (y_j)_{j \in J}$ for the collection of all production rules, etc.

A social choice function is implementable as a Bayes-Nash equilibrium if there is a game with Bayes-Nash equilibrium, so that the equilibrium allocation of this game coincides in each state of the economy with the allocation stipulated by the social choice function. For the given setup, the revelation principle holds, so that we can without loss of generality limit attention to the implementation of a social choice function via a direct mechanism that induces a game in which truth-telling is a Bayes-Nash equilibrium. Thus, we say that a social choice function is incentive-compatible if truth-telling is a Bayes-Nash equilibrium of the corresponding direct mechanism.

Incentive-compatibility. Incentive-compatibility for consumer i holds, provided that for each $\theta_i^l \in \Theta_i$ and for all $\theta_i^k \in \Theta_i$,

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^k \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^k)) - T(\theta_i^k) \ , \tag{IC_C}$$

where $V(\theta_i^k \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^k)) := \mathbb{E}_{(\theta_{-i},\delta)}\left[v(\theta_i^l, q_i(\theta_{-i}, \theta_i^k, \delta))\right]$ is the expected consumption utility for type θ_i^l of consumer i in case of announcing θ_i^k to the mechanism designer, given that all other consumers and producers reveal their preferences and technologies. Analogously, $T(\theta_i^k) := \mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\theta_{-i}, \theta_i^k, \delta)\right]$ is i's expected payment in case of reporting a preference parameter θ_i^k. The expectations operator $\mathbb{E}_{(\theta_{-i},\delta)}$ indicates that expectations are computed with respect to the random variable (θ_{-i}, δ). By contrast, the realization of θ_i is known when computing this expectation.

Likewise, incentive-compatibility for firm j requires that for all $\delta_j^l \in \Delta_j$ and for all $\delta_j^k \in \Delta_j$,

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^k) - K(\delta_j^k \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^k)) \ , \tag{IC_F}$$

where $R(\delta_j^k) := \mathbb{E}_{(\theta,\delta_{-j})}\left[r_j(\theta, \delta_{-j}, \delta_j^k)\right]$ is j's expected revenue in case of reporting a cost parameter δ_j^k, and $K(\delta_j^k \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^k)) := \mathbb{E}_{(\theta,\delta_{-j})}[k(\delta^l, y_j(\theta, \delta_{-j}, \delta_j^k))]$ is the expected cost for type δ_j^l of firm j in case of announcing δ_j^k to the mechanism designer.

Participation Constraints. Social choice functions have to respect lower bounds on the consumers' utility and the producers' profits, respectively. Formally, we require that for all i and for all $\theta_i^l \in \Theta_i$,

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq \underline{u}_i \ , \tag{PC_C}$$

where $\underline{u}_i$ denotes a lower bound for the expected utility of consumer i. Likewise, for all j and $\delta_j^l \in \Delta_j$,

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \underline{\pi}_j \ , \tag{PC_F}$$

where $\underline{\pi}_j$ is a lower bound for the expected profit of firm j.

The interpretation of these participation constraints depends on the application at hand. For instance, we may think that the implementation of the given social choice function replaces a status quo outcome and moreover requires a unanimous consent of all consumers and producers. In this case, $\underline{u}_i$ and $\underline{\pi}_j$ would, respectively, be interpreted as consumer i's and producer j's payoff in the status quo. Alternatively, in a model in which a government has coercive power, such a consent may not be needed but producers may have the possibility to shut down, so that a social choice function has to provide them at least with the level of profits that they would realize in this case. By choosing $\underline{u}_i$ and $\underline{\pi}_j$ arbitrarily small, we can also capture situations for which participation constraints are irrelevant.

Physical constraints. For many applications we assume that the consumers' consumption is bounded by the total output that is made available by the producers. Denote total output by $Y(\theta, \delta) = \sum_{j=1}^m y_j(\theta, \delta)$. If we consider an allocation problem involving private goods, then it has to be the case that, for all (δ, θ), $\sum_{i=1}^n q_i(\theta, \delta) \leq Y(\theta, \delta)$. If the good is non-rival and non-excludable then, for all i and all (δ, θ), $q_i(\delta, \theta) = Y(\delta, \theta)$. If the good is non-rival, but excludable, then, for all i and all (δ, θ), $0 \leq q_i(\delta, \theta) \leq Y(\delta, \theta)$. We capture all these cases by postulating that, for all (θ, δ),

$$(q_i(\theta, \delta))_{i \in I} \in \Lambda(Y(\delta, \theta)) \ , \tag{1.5}$$

where $\Lambda(Y(\delta, \theta))$ is an abstract consumption set. Its structure depends on whether the goods in question are public or private.

Budget balance. We often assume that a social choice function has to satisfy a budget constraint, which requires that the consumers' expected payments suffice to cover the producers' expected revenues,

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^n t_i(\theta, \delta)\right] \geq \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^m r_j(\theta, \delta)\right] \ . \tag{1.6}$$

An alternative that we will also consider is that the consumer's expected payments have to be sufficient to cover the producer's expected costs, i.e.,

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] \geq \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m} k(\delta_j, y_j(\theta,\delta))\right] . \tag{1.7}$$

The budget condition in (1.6) is relevant in models in which producers have private information. The budget condition in (1.7) is employed in models in which the producers' cost functions are assumed to be publicly known information and in which profits in the hands of producers are considered undesirable. Since there is no private information there is also no impediment to reaching an outcome with zero expected profits, i.e., with

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m} r_j(\theta,\delta)\right] = \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m} k(\delta_j, y_j(\theta,\delta))\right] .$$

However, as we will see below, such an outcome is out of reach if producers have private information and if their participation in the system is voluntary so that $\pi_j(\delta_j, r_j, y_j) \geq \underline{\pi}_j$, for all j.

These budget conditions allow for the possibility that there are deficits in some states of the economy and surpluses in others, provided that, in expectation, the surpluses are at least as large as the deficits. Thus, it is more permissive than having a separate budget balance condition for each state of the economy. There are various justifications for working with this permissive notion of budget balance. First, for many applications of the independent private values model, the following proposition holds true: If there is a social choice function that is incentive-compatible, respects the relevant participation constraints and budget balance in expectation, there is an 'equivalent' social choice function that satisfies in addition a state-wise requirement of budget balance, see Börgers and Norman (2009). Second, a requirement of budget balance in expectation may be justified with an appeal to the Law of Large Numbers.[5] If the numbers of consumers and producers is large, the discrepancy between budget balance in expectation and budget balance for each state separately becomes small, see Bierbrauer (2011b). Finally, many analyses of the independent private values model have established impossibility results, see Myerson and Satterthwaite (1983) or Mailath and Postlewaite (1990). If there is no social choice function that satisfies budget balance in expectation, then there is also no social choice function that gives rise to budget balance in each state separately. Thus, for the purpose of establishing an impossibility result, working with the requirement of budget balance in expectation can be a useful modeling device.

[5] See e.g. Judd (1985) and Feldman and Gilles (1985).

Surplus measures. The total expected surplus that is generated by a social choice function is given by

$$S((q_i)_{i\in I},(y_j)_{j\in J}) = \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n} v(\theta_i, q_i(\theta,\delta)) - \sum_{j=1}^{m} k(\delta_j, y_j(\theta,\delta))\right] .$$

In a model with quasi-linear preferences, a social choice function is Pareto efficient if and only if the relevant budget constraint holds as an equality, the participation constraints in (PC_C) and (PC_F) are satisfied and $(q_i)_{i\in I}$, and $(y_j)_{j\in J}$ are chosen so as to maximize total surplus $S((q_i)_{i\in I},(y_j)_{j\in J})$ subject to the constraint of physical feasibility in (1.5). Note that there are typically many different Pareto efficient social choice functions. While the criterion of surplus-maximization pins down the functions $(q_i)_{i\in I}$ and $(y_j)_{j\in J}$, alternative specifications of the payment and revenue rules $(t_i)_{i\in I}$ and $(r_j)_{j\in J}$ give rise to different distributions of the surplus among consumers and producers.

1.5 Implementable provision rules

Before we turn to the question under which conditions efficient outcomes can be obtained, we will provide, as a preliminary step, a characterization of the set of implementable social choice functions, i.e., social choice functions with the property that there exists a direct mechanism that is incentive compatible, satisfies participation constraints, and is budgetary and physically feasible.

We begin by deriving a necessary and a sufficient condition for a social choice function to be implementable as Bayes-Nash equilibrium.

1.5.1 Necessary condition

The following proposition states a necessary condition for the possibility to implement a social choice function. More specifically, it states an inequality constraint so that, if this inequality is violated, we know that there is no mechanism that satisfies the incentive compatibility constraints in (IC_C) and (IC_F), participation constraints in (PC_C) and (PC_F), and the expected budget constraint in (1.6).

Proposition 1.3. $\left\{(q_i)_{i=1}^n,(y_j)_{j=1}^m\right\}$ *is part of an implementable social choice function only if*

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] - \sum_{i=1}^{n}\underline{u}_i \geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j,y_j(\theta,\delta)) + \{k(\delta_j,y_j(\theta,\delta)) - k(\delta_j^-,y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right] + \sum_{j=1}^{m}\underline{\pi}_j .$$

A proof of the Proposition is in part 1.B of the Appendix. The social choice function can be split into a transfer and consumption rule for consumers and a revenue and production rule for producers. We begin by holding fixed consumers' consumption rule and derive the maximal transfers that consumers are able to make if incentive compatibility and participation constraints need to be both respected at the same time. Similarly, we hold fixed firms' production rule and derive the minimal revenues that firms are willing to accept if again incentive compatibility and participation constraints need to be respected. We then check whether consumers maximal transfers cover firms' minimal revenues. If the mechanism is unable to generate consumer payments that are high enough to cover the minimal revenue of provision, then there is no mechanism that reaches efficiency.

We make use of techniques used in the non-linear pricing literature (see e.g. Mussa and Rosen (1978)). We consider the *relaxed problem* of maximizing consumers' transfers subject to the *local downward* incentive compatibility constraints and the participation constraints for the lowest preference type. The local downward incentive compatibility constraints prevent type θ_i^l of consumer i to announce the next lower type to the mechanism designer. Hence, only a subset of consumers' incentive compatibility and participation constraints are taken into account. Thus, the maximal transfers that can be obtained at the solution to the relaxed problem, is an upper bound on the transfers that can be obtained if all incentive compatibility constraints are taken into account. We then show that the maximal transfers are given by the left-hand-side of the inequality constraint in Proposition 1.3 above.

Similarly, we solve a relaxed problem on the production side, to define the minimal revenues firms need to receive, where the *local upward* incentive compatibility constraints and the participation constraints of the worst technology type δ^r are taken into consideration. As under consumer transfer maximization, only a subset of firms' incentive compatibility constraints is considered, so that the minimal revenues that need to be generated at the solution to the relaxed problem, is a lower bound on the revenues that can be generated if all incentive compatibility constraints are taken into account. If we plug the maximal transfers and the minimal revenues in the requirement of the budget constraint in (1.6), we obtain the inequality in Proposition 1.3 above.

For the derivation of the necessary condition in Proposition 1.3, we did not assume that the utility function u_i and the profit function π_j are differentiable. Neither did we need to assume, that q_i and y_j are monotonic. In contrast to a model that assumes the set of possible types for all agents to be infinite, we can avoid such assumptions on endogenous objects.

1.5.2 Sufficient condition

The next proposition establishes a sufficient condition for the implementability of a social choice function. For this purpose, we consider only a subset of all provision and production rules, namely those that satisfy the following monotonicity conditions: For every i, for all (θ_{-i}, δ) and all $l \in \{0, s-1\}$,

$$q(\theta_i^+, \theta_{-i}, \delta) \geq q(\theta_i, \theta_{-i}, \delta) \ , \tag{1.8}$$

i.e., consumption must be monotonically increasing in θ_i for every consumer i. For every j, for all (θ, δ_{-j}) and all $l \in \{2, ..., r\}$,

$$y(\delta_j^-, \delta_{-j}, \theta) \geq y(\delta_j, \delta_{-j}, \theta) \ , \tag{1.9}$$

i.e., firm j's contribution to production must be monotonically decreasing in δ_j.

Efficient mechanisms satisfy these monotonicity conditions. Thus, as long as we limit ourselves to surplus-maximizing provision rules, the assumptions on monotonicity are not restrictive.

Proposition 1.4. *Let the monotonicity constraints in (1.8) and (1.9) be satisfied, then we can implement the social choice function if*

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] - \sum_{i=1}^{n}\underline{u}_i \geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j(\theta,\delta)) + \{k(\delta_j, y_j(\theta,\delta)) - k(\delta_j^-, y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right] + \sum_{j=1}^{m}\underline{\pi}_j \ .$$

Suppose that the condition in Proposition 1.4 holds. We need to show that we can construct a payment scheme that satisfies all relevant constraints. We can choose our transfer and revenue scheme such that they solve the relaxed problems that we studied in the proof of Proposition 1.3. After that, we need to verify that the transfer and revenue schemes, which solve the relaxed problems, satisfy not only the local downward incentive compatibility constraints of consumers and the local upward incentive compatibility constraints of firms, but all consumers' and firms' incentive compatibility constraints. We prove that the fact that all local downward incentive compatibility constraints of consumers are binding, together with the monotonicity constraints of consumption, implies that all incentive compatibility constraints are satisfied. With that, consumption is efficient for the consumer of highest type and distorted downwards for all other consumers. And similarly, production is efficient for the firm of lowest costs and distorted upwards for all other firms.

Definition 1.1. *We define the minimal subsidy* $(MS(\cdot))$ *as*

$$\begin{aligned} MS((q_i)_{i\in I},(y_j)_{j\in J}) &:= \sum_{i=1}^{n} \underline{u}_i - \sum_{j=1}^{m} \underline{\pi}_j \\ &\left.- \mathbb{E}_{(\theta,\delta)} \left[\sum_{i=1}^{n} v(\theta_i, q_i(\theta,\delta)) - \sum_{j=1}^{m} k(\delta_j, y_j(\theta,\delta)) \right] \right\} S(\cdot) \\ &\left.\begin{aligned} &+ \mathbb{E}_{(\theta,\delta)} \left[\sum_{i=1}^{n} \{ v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta)) \} \frac{1-F(\theta_i)}{f(\theta_i)} \right. \\ &\qquad \left. - \sum_{j=1}^{m} \{ k(\delta_j, y_j(\theta,\delta)) - k(\delta_j^-, y_j(\theta,\delta)) \} \frac{P(\delta_j)}{p(\delta_j)} \right] \end{aligned} \right\} IR(\cdot) \end{aligned}$$

Under the monotonicity constraints (1.8) and (1.9), the minimal subsidy gives the amount of money that is required from an outside party to satisfy Proposition 1.4.[6] Apart from the reservation utilities and profits, it consists of two components: The first component is surplus $S(\cdot)$, given by the difference of consumers' valuations and firms' costs. The second component $IR(\cdot)$ is the difference of consumers' and firms' information rents. In an environment with complete information, the maximal transfer a consumer i is is willing to pay is his true valuation. With private information about preferences, however, the transfers are decreased by the expression $\{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}$. That is why these expressions are often interpreted as information rents. Similarly, for firms, the virtual costs account for the private information. The revenues, firms would accept in an environment with complete information, are increased by an information rent.

1.5.3 Efficiency

In the previous section, we have been concerned with deriving conditions such that social choice functions could be implemented in an environment where agents' have private information about their preferences and costs, respectively. Now, we focus on the welfare evaluation of implementable social choice functions.

[6] This notion follows Myerson and Satterthwaite (1983), who call the amount that would be required from an outside party to overcome the impossibility result *minimal lump-sum subsidy.*

Corollary 1.1. *A mechanism that maximizes surplus $S(\theta,\delta)$, in the following denoted by $\{(q_i^*)_{i=1}^n,(y_j^*)_{j=1}^m\}$, is implementable if and only if*

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^n\left(v(\theta_i,q_i^*(\theta,\delta))-\{v(\theta_i^+,q_i^*(\theta,\delta))-v(\theta_i,q_i^*(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right]-\sum_{i=1}^n\underline{u}_i\geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^m\left(k(\delta_j,y_j^*(\theta,\delta))+\{(\delta_j,y_j^*(\theta,\delta))-k(\delta_j^-,y_j^*(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right]+\sum_{j=1}^m\underline{\pi}_j\ .$$

A necessary condition for the implementability of $((q_i^*)_{i=1}^n,(y_j^*)_{j=1}^m)$ is that the maximal transfers that can be extracted from consumers suffice to cover the minimal revenues that the production sector needs to obtain, which means that the minimal subsidy is negative.

Sufficiency can be shown by constructing the transfers of consumers such that all local downward incentive compatibility constraints are binding, and constructing the revenues of producers such that all local upward incentive compatibility constraints are binding. Finally, we choose the transfer of the consumer with the lowest preference parameter such that budget balance is satisfied and the minimal utilities and resource requirement is taken into consideration. Then, we obtain a mechanism that achieves $((q_i^*)_{i=1}^n,(y_j^*)_{j=1}^m)$ and satisfies all relevant constraints stated in Corollary 1.1.

From Corollary 1.1, it follows immediately that when participation constraints can be violated, a social choice function can be implemented efficiently. The minimal subsidy can be decreased without limit. This is an alternative proof of the possibility results obtained by D'Aspremont and Gerard-Varet (1979) and Arrow (1979).

The following Corollary shows how the second-best mechanism can be derived.

Corollary 1.2. *The second-best mechanism, in the following denoted by $((q_i^{**})_{i=1}^n,(y_j^{**})_{j=1}^m)$, can be found by maximizing $S(\theta,\delta)$ subject to*

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^n\left(v(\theta_i,q_i(\theta,\delta))-\{v(\theta_i^+,q_i(\theta,\delta))-v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right]-\sum_{i=1}^n\underline{u}_i\geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^m\left(k(\delta_j,y_j(\theta,\delta))+\{(\delta_j,y_j(\theta,\delta))-k(\delta_j^-,y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right]+\sum_{j=1}^m\underline{\pi}_j$$

and the monotonicity constraints in (1.8) and (1.9).

In the solution to the second-best problem, the budget condition has to be binding. Otherwise, expected consumer transfers can be reduced without violating any of the incentive compatibility and participation constraints. The provision rule that solves the second-best problem satisfies the monotonicity constraints for all consumers and all firms. This follows because the optimal provision level is given by the first-order condition where the sum of virtual marginal valuation equals the sum of virtual marginal costs. The assumption that hazard rates are non-decreasing for consumers implies that virtual valuation is increasing in consumer i's type, and the assumption that the monotone hazard rate is non-increasing for firms implies that virtual costs are increasing in the cost type.

Continuous types. To study whether a model with a large but discrete number of types behaves approximately in the same way as a model with continuum types, we state the expression for the minimal subsidy for continuous type distributions. We know from the literature (compare e.g. Mas-Colell, Whinston, and Greene, 1995) that for a surplus maximizing mechanism $\left\{(q_i^*)_{i=1}^n, (y_j^*)_{j=1}^m\right\}$ in a continuous environment, implementability is possible if the constraint in the following Remark is satisfied. It is based on the assumption that functions are continuously differentiable.

Remark 1. *Suppose that q_i and y_j are continuously differentiable functions. Then a mechanism that maximizes $S(\theta, \delta)$, is implementable if and only if*

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i, q_i^*(\theta,\delta)) - v_1(\theta_i, q_i^*(\theta,\delta))\frac{1-F(\theta_i)}{f(\theta_i)}d\theta_i\right)\right] - \sum_{i=1}^{n}\underline{u}_i \geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j^*(\theta,\delta)) + k_1(\delta_j, y_j^*(\theta,\delta))\frac{P(\delta_j)}{p(\delta_j)}d\delta_j\right)\right] + \sum_{j=1}^{m}\underline{\pi}_j. \tag{1.10}$$

The differences, in comparison to Corollary 1.1, are that $\{v(\theta_i^+, q_i^*(\theta,\delta)) - v(\theta_i, q_i^*(\theta,\delta))\}$ and $\{k(\delta_j, y_j^*(\theta,\delta)) - k(\delta_j^-, y_j^*(\theta,\delta))\}$ are replaced by the derivatives $v_1(\theta_i, q_i^*(\theta,\delta))$ and $k_1(\delta_j, y_j^*(\theta,\delta))$, respectively. Therefore, the differentiability assumption is crucial with a continuum of types.

1.6 Comparative statics I: From few to many agents

We use our results in Proposition 1.3 and 1.4 to obtain a more systematic understanding of how a change in the parameters of the model affects the possibility to implement efficient social choice functions. Again, we will check whether our results depend on whether the allocation problem involves private or public goods.

1.6.1 Possibility results when the type set is binary

Bilateral trade. We have shown on Section 1.3.1 that parameter constellations can be found such that efficient bilateral trade is possible. In order to investigate what drives this possibility result, we make some simplifying assumptions.

Assumption 1.1. *All agents have a binary type set.*[7]

Under Assumption 1.1, the buyer's marginal valuation can be high or low, $\Theta = \{\theta^L, \theta^H\}$, and the seller's marginal costs can take two values, $\Delta = \{\delta^L, \delta^H\}$. We denote the respective probabilities by $Prob(\theta = \theta_L) = f^L$, $Prob(\theta = \theta^H) = f^H = 1 - f^L$ and $Prob(\delta = \delta^L) = p^L$, $Prob(\delta = \delta^H) = p^H = 1 - p^L$.

Assumption 1.2. ***(Symmetry)*** *The distance between the low type and the high type is the same for buyer and seller, $\theta^H - \theta^L = \delta^H - \delta^L$.*[8] *The probability of a high valuation buyer is equal to the probability of a low cost seller $f^H = p^L$ and therefore $f^L = p^H$.*

Analogously to the continuous environment of Myerson and Satterthwaite (1983), where $\left(interior[\theta^L, \theta^H] \cap interior\left([\delta^L, \delta^H]\right) \neq \emptyset\right.$ and $[\theta^L, \theta^H] \neq [\delta^L, \delta^H]$, we assume that the parameter constellation is such that $\delta^L < \theta^L < \delta^H < \theta^H$, with the normalization $\delta^L = 0$ and $\theta^H = 1$. The social choice function f is efficient if and only if $q(\theta, \delta) = y(\theta, \delta)$, $q(\theta, \delta)$ satisfies

$$q^f(\theta, \delta) = \begin{cases} 0, & \text{if} \quad \theta < \delta \ , \\ \in [0, 1] & \text{if} \quad \theta = \delta \ , \\ 1, & \text{if} \quad \theta > \delta \ , \end{cases}$$

and $\mathbb{E}_{(\theta,\delta)}[t^f(\theta, \delta)] = \mathbb{E}_{(\theta,\delta)}[r^f(\theta, \delta)]$.

The following Observation follows from Matsuo (1989) under Assumptions 1.1 and 1.2.

Observation 1.1. *Suppose Assumptions 1.1 and 1.2 are satisfied and consider the direct mechanism for f. There is a social choice function, which is efficient, implementable as Bayes-Nash equilibrium, and yields a non-negative material payoff for every type of agent if and only if*

$$f^L(\theta^H - \delta^L) + (\theta^L - \delta^H) > 0 \quad \Leftrightarrow \quad f^L > \delta^H - \theta^L \ .$$

The proof of this Observation and all following Observations can be found in Appendix 1.E. It provides a sufficient condition for the possibility to reach efficient bilateral trade. Specifically, the condition states that the probability of the low valuation buyer, respectively on the high cost seller, needs to be bigger than the ratio that measures the overlap of buyer's and seller's type sets $(\delta^H - \theta^L)$, relative to the whole type set that we normalized to 1 $(\theta^H - \delta^L = 1)$. We call this ratio in the following d,

$$d := \frac{\delta^H - \theta^L}{\theta^H - \delta^L} = \delta^H - \theta^L \ .$$

[7]Without loss of generality, we specify the binary type set, in the following, as consisting of a low and high type, where the high type exceeds the low type.

[8]We assume that $\Theta \subseteq \mathbb{R}$ and $\Delta \subseteq \mathbb{R}$, such that 'distance' is well-defined by the Euclidean metric.

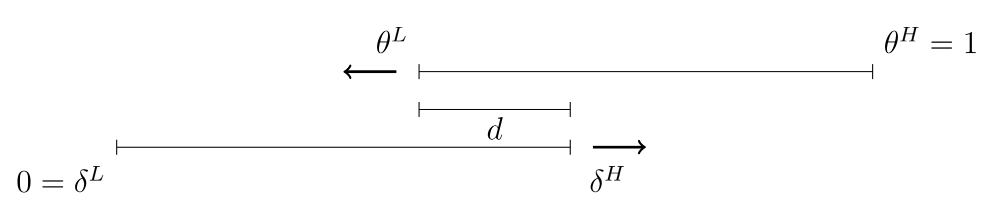
The bigger the ratio d, the bigger is the overlap of buyer's and seller's type sets. If $d = 0$, the type sets of buyer and seller are disjunct, and efficient trade is always possible, regardless of the probability distribution. On the other hand, the maximum of d is given by 1, where the type sets of the buyer and the seller are congruent.

To understand what drives the possibility result above, it is instructive to look at the proof of Observation 1.1. By Proposition 1.4, we know that efficient bilateral trade is possible if the minimal subsidy is negative. Given that trade is efficient for all states of the economy but when the low valuation buyer faces the high cost seller, we have three states of the economy that need to be evaluated, and weighted with the probability of occurrence.

$$\text{MS} = -\left(\underbrace{\begin{array}{l} \; f^H f^L(\theta^H - \delta^H) \\ +\; f^H f^H(\theta^H - \delta^L) \\ +\; f^L f^H(\theta^L - \delta^L) \end{array}}_{S(\cdot)} \begin{array}{c} - \\ \\ - \end{array} \underbrace{\begin{array}{l} f^H f^L(\delta^H - \delta^L)\frac{f^H}{f^L} \\ \\ f^L f^H(\theta^H - \theta^L)\frac{f^H}{f^L} \end{array}}_{IR(\cdot)} \right)$$

In the following we do comparative statics for different components of the model. As a first exercise, we fix the probability of the low valuation buyer and the high cost seller f^L and vary the ratio d. Without loss of generality, we keep the normalization of the whole type set, so that $\theta^H - \delta^L = 1$. In order to change d, the position of θ^L and δ^H can be varied. By Assumption 1.2, $\theta^H - \theta^L = \delta^H - \delta^L$, so that a shift from the low valuation buyer to the left goes hand in hand with a move of the high cost seller to the right, and vice versa, as the following Figure illustrates.

Figure 1.2: Changing parameters in the bilateral trade setting



Observation 1.2. *Suppose we move the types as in Figure 1.2 above, such that d is increased. This affects the minimal subsidy via the expected information rents and the expected surplus, where*

$$\frac{\partial IR(\cdot)}{\partial d} > 0 , \qquad \textit{and} \qquad \frac{\partial S(\cdot)}{\partial d} < 0 .$$

For types outside the overlap, efficient trade takes place – independent of the opponent's type, i.e., the private information of the other party does not matter for the decision whether trade takes place. This means, that the two-sided private information only matters in the overlap. Thus, d can be understood as a measure for the importance of two-sided private information. The first effect depends on the distance of types for each player. If d increases, higher expected information rents need to be paid under the new parameter constellation, since the distance of high and low types increases for both players. This has a negative effect on the possibility to achieve efficient trade, and therefore yields a negative information rent effect. Second, if d increases, the expected surplus decreases, i.e., we have a negative surplus effect. Further, the average valuation for the good goes down relative to the average costs if d is increased. Therefore, expected surplus decreases, which has a negative effect on the possibility to achieve efficient trade. Thus, an increase of d decreases expected surplus and increases expected information rents. It is hence *more costly* to achieve efficiency.[9]

Next, consider the situation where for a given d and f^L implementation is possible. We fix the ratio that measures the overlap d and analyze how varying the probability f^L affects implementability.

Observation 1.3. *Suppose d is fixed and the probability f^L is increased. This affects the minimal subsidy via the expected information rents and the expected surplus, where*

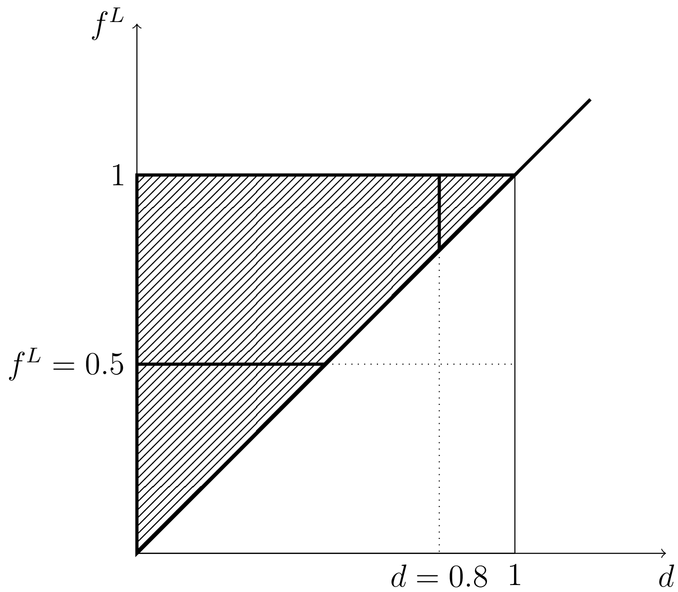
$$\frac{\partial IR(\cdot)}{\partial f^L} < 0 , \qquad \textit{and} \qquad \frac{\partial S(\cdot)}{\partial f^L} < 0 .$$

Incentive compatibility constraints are binding for the high valuation buyer and the low cost seller, and slack for the other types. Hence, if f^L increases, we will have to pay less individuals information rents. This has a positive effect on the possibility to achieve efficient trade. Second, we find that the surplus effect is negative if $\theta^H - \delta^H = \theta^L - \delta^L < \frac{1}{2}$. Since we start our analysis in a situation where implementation is possible, we know that this conditions always holds. Thus, the increase of f^L has an unambiguous negative effect on implementability since it lowers the expected surplus. Intuitively, the

[9]Whenever we write efficient implementation gets more costly, we mean the following: If we compare two sets that represent the tuples of f^L and d, for which efficient trade is possible the set for which efficiency is more costly is a subset of the other. Analogously for the case where efficiency is less costly to achieve.

expected surplus decreases since we expect more low valuation buyers, respectively high cost sellers. Overall, an increases in f^L increases the minimal subsidy, and thus has a negative effect on the possibility to reach efficient trade: By Observation 1.1, we know that an increase of f^L makes it less costly to achieve efficient trade. Therefore, when f^L is increased, the reduction in expected information rents is bigger than the reduction in expected surplus, so that an increase of f^L makes it less costly to achieve efficiency. The intuition for this result is that there is more mass on the types that do not receive an information rent. Overall, the effect of an increase in f^L on the minimal subsidy is negative such that implementation gets less costly.

Figure 1.3: Possibility of efficient trade



By assumption, f^L and d are bounded by 0 and 1. If f^L is for example fixed at 0.5, then the distance between the high cost seller and the low valuation buyer needs to be lower than 0.5 in order to achieve efficient trade. And if the distance is for example fixed at $d = 0.8$, then the probability of the low type buyer, respectively the high cost seller, needs to be higher than 0.8 to get a possibility result.

Figure 1.3 illustrates the interplay of the expected surplus effect and the expected information rent effect on the possibility of efficient bilateral trade. The grey shaded area T_0 gives every combination of f^L and d, such that, efficient trade is possible,

$$T_0 = \left\{ (f^L, d) : f^L > d = \frac{\delta^H - \theta^L}{\theta^H - \delta^L} \right\} .$$

Public Good. Given Assumption 1.1 holds, a social choice function f' specifies whether the public good is provided $q^{f'}(\theta_1, \theta_2)$, and the accompanying transfers $t_1^{f'}(\theta_1, \theta_2)$ and $t_2^{f'}(\theta_1, \theta_2)$. Pareto efficient public good provision requires

$$q^{f'}(\theta_1, \theta_2) = \begin{cases} 0, & \text{if } \frac{1}{n}\sum_{i=1}^{n} \theta_i < c\ , \\ \in \{0, 1\} & \text{if } \frac{1}{n}\sum_{i=1}^{n} \theta_i = c\ , \\ 1, & \text{if } \frac{1}{n}\sum_{i=1}^{n} \theta_i > c\ . \end{cases}$$

and $\mathbb{E}_{(\theta)}[t_1^{f'}(\theta_1, \theta_2) + t_2^{f'}(\theta_1, \theta_2)] = 2c\, \mathbb{E}_{(\theta)}[q^{f'}(\theta)]$ for all $(\theta_1, \theta_2) \in \Theta$.

The following Observation highlights that if there are just two consumers, then efficient implementation of the public good will be possible for some parameter constellations.

Observation 1.4. *Suppose Assumption 1.1 is satisfied, the per capita costs are such that $\theta^L < c < \theta^H$, and consider the direct mechanism for f'. There is a social choice function which is efficient, implementable as Bayes Nash equilibrium, and it yields non-negative utility for every type of consumer if and only if*

$$f^L - \frac{c - \theta^L}{\theta^H - c} > 0\ .$$

Observation 1.4 provides a sufficient condition for efficient public good provision. Following the proof of Observation 1.4, the public good can be provided efficiently, if and only if consumers' expected transfers cover the costs of public good provision, so that the minimal subsidy is negative

$$MS = -\left(\underbrace{\left[\theta^H - c - f_L(c - \theta^L)\right]}_{S(\cdot)} - \underbrace{(1 - f^L)(\theta^H - \theta^L)}_{IR(\cdot)} \right) < 0\ .$$

The first term denotes expected surplus, depending on the cost c and the type distribution, whereas the second term describes expected information rents.

In the following, we analyze what drives the possibility to reach efficient public good provision. Again, we analyze the effect of a change in the type distribution f^L and a change in the cost parameters c and split the resulting effects into the expected surplus and the expected information rent effect.

Observation 1.5. *Suppose Assumptions 1.1 and 1.2 hold.*

i) Suppose further that f^L is fixed, then $\frac{\partial S(\cdot)}{\partial c} < 0$.

ii) Suppose further that c is fixed, then

$$\frac{\partial IR(\cdot)}{\partial f^L} < 0\ , \qquad \textit{and} \qquad \frac{\partial S(\cdot)}{\partial f^L} < 0\ .$$

When costs are increased, this does not affect the information rents of the agents, but lowers the expected surplus. Higher per capita costs imply that in expectation, the share of benefiting types reduces compared to the share of suffering types. This means that for higher c, the average valuation for the public good decreases compared to the fixed per capita costs.

If c is fixed, a change in f^L affects the minimal subsidy in two ways: First, the expected surplus decreases if f^L increases. This has a negative effect on the possibility of efficient public good provision. And second, for the expected information rents, we find, that if f^L increases, we have to pay in expectation less agents an information rent. This means, the higher the probability for low valuation consumers, the lower the expected information rent term and the higher the possibility for efficient public good provision, ceteris paribus. As we can see in Observation 1.4, the reduced expected information rent effect dominates the negative expected surplus effect, such that an increase in f^L makes it less costly to provide the public good efficiently, i.e., the minimal subsidy decreases.

Comparison. When a binary type set and two agents are considered, efficient bilateral trade and efficient public good provision are possible when parameters are chosen appropriately. Further, when the average valuation for the private or public good is increased, it gets cheaper to achieve efficiency, in the sense that the minimal subsidy goes down. To understand the difference between private and public goods, we study how the increase in the number of the agents affects the minimal subsidy.

1.6.2 Many agents

The classical bilateral trade setting has one buyer and one seller. To that extent Observation 1.1 is not restrictive. Contrary, the impossibility result in the public good setting is studied for a large economy, with many individuals. In the following, we want to analyze these impossibility results under the assumption that the type set of individuals remains binary.

Public good. Mailath and Postlewaite (1990) show that efficient public good provision is impossible when the number of individuals goes to infinity. The following Proposition shows that the assumption of continuous type sets has no impact on this impossibility result. Even for a binary type set, efficient public good provision is impossible if the number of individuals grows without limit.

Proposition 1.5. *Suppose Assumption 1.1 holds. Consider the public good example and an economy with n individuals. For any sequence of incentive-compatible mechanisms $(q, t_1^n, \ldots, t_n^n)$,*

$$lim_{n \mapsto \infty} \mathbb{E}_{(\theta)} \left[\frac{1}{n} \sum_{i=1}^{n} t_i^n(\theta) \right] = 0 \, .$$

As $n \to \infty$, the probability of public good provision converges to 0 if $c > \theta^L$, and converges to 1 otherwise.

According to Proposition 1.5, the per capita revenue from individuals' contribution goes to zero. That is, with many individuals, any social choice function, which is attainable when voluntary participation needs to be guaranteed, prescribes a public good provision level that is equal to 0. Even if the average valuation for the public good is larger than the per capita costs, the amount that is implemented is almost equal to 0, although the efficient amount is almost equal to 1.

The reason for the impossibility result is that, for $n \to \infty$, any individual's impact on the public good provision becomes negligible. The free-rider problem in public good provision becomes extreme as the number of individuals becomes large. Since no individual is pivotal for the production of the public good, incentive compatibility implies that all individuals have to make the same lump-sum transfer that does not depend on their announced preference intensity. Participation constraints imply that this lump-sum transfer must not exceed θ^L. Thus, the aggregate of all individuals' transfers is as if every individual has a low valuation for the public good. This makes it impossible to cover the costs of public good provision. Hence, for the impossibility result, the assumption that the type set is discrete has no impact. Remember that Proposition 1.2 states that public good provision is as well not possible under the assumption of a continuous type set when the number of individuals goes to infinity.

Private good. Consider a finite economy with $I = \{1, \dots n\}$ buyers and $J = \{1, \dots n\}$ sellers. Buyers and sellers have equivalent binary type sets with the corresponding probability distributions (f_i^L, f_i^H) and (p_j^L, p_j^H).[10] The agents face a price $\rho \in [\theta_i^L, \delta_i^H]$. As we have seen in previous sections, there exist parameter constellations and probability distributions, where trade does not take place. Now, we assume that the number of agents grows without limits, i.e., $n \longrightarrow \infty$. The Law of Large Numbers applies, and probabilities can be interpreted as cross-sectional distribution of types. In particular, this means that for large economies one knows with probability 1, that there exist high type buyers with a share of f_i^H, and low type sellers, with a share of $p_j^L = f_i^H$. For these agents, trade always takes place. This proves the existence of trade in large economies for prices $\rho \in [\theta_i^L, \delta_i^H]$ if agents have binary type sets.[11]

Gresik and Satterhwaite (1989) additionally show that trade in large economies is efficient – ex-ante and ex-post – holding the ratio of buyer and seller fixed. Further, they provide results on the rate of convergence to the competitive equilibrium.

Comparison. When private and public goods are compared for economies with many agents, we see that efficient public good provision is impossible when the number of con-

[10] $f_i^H, f_i^L, p_j^H, p_j^L > 0, \forall\, i, j$.

[11] This argument holds independently of whether the type set is finite or infinite.

sumers is large. Contrary, for the private good model efficient bilateral trade is possible if there are many agents. The impossibility result of Mailath and Postlewaite (1990) is thus *stronger* than the impossibility result of Myerson and Satterthwaite (1983) in the sense that it extends to any model with a discrete set of types.

1.7 Comparative Statics II: From few to many types

This section shows that the minimal subsidy for a discrete type set with a large number of types converges to the minimal subsidy for a continuous type set environment. We proceed as follows: We first show how one additional type affects the implementation rule for private and public goods and add a third type for each agent. To study the changes of an increased number of types separately from changes in the number of agents, we hold the number of agents fixed in this section. Second, we consider the situation where the number of types gets larger and larger and study general convergence. We provide conditions that need to be met for the convergence result and discuss how violating these conditions affects our analysis.

1.7.1 Introducing a third type.

Bilateral trade. In the following, we consider what happens if we add step-by-step new types in *rounds* to agents' type sets of the pre-round. Thereby, we do not move the existing types, but add new types 'between' them.[12] If the procedure of adding types fulfills two conditions we call it 'uniform extension'.

Definition 1.2. (Uniform Extension) *Consider the agent sets I and J, fully ordered finite type sets $\Theta_{0i} = \{\theta_{0i}^L, \dots, \theta_{0i}^H\},\ \#\Theta_{0i} < \infty, \forall\, i \in I, \Delta_{0j} = \{\delta_{0j}^L, \dots, \delta_{0j}^H\},\ \#\Delta_{0j} < \infty, \forall\, j \in J$ and the respective probabilities, such that $f_{0i}^L + \dots + f_{0i}^H = 1$ and $p_{0j}^L + \dots + p_{0j}^H = 1$. We add finitely many new types in every round k to every agent's type set, such that for every new type θ_{ki} and δ_{kj}, it holds that $\theta_{0i}^L < \theta_{ki} < \theta_{0i}^H$ and $\delta_{0j}^L, < \delta_{kj} < \delta_{0j}^H$, respectively. We call this procedure uniform extension if it has the following properties*

- $\overline{\{\theta_{0i}^L, \dots, \theta_{0i}^H\}} = [\theta_{0i}^L, \theta_{0i}^H]$ *and* $\overline{\{\delta_{0j}^L, \dots, \delta_{0j}^H\}} = [\delta_{0j}^L, \delta_{0j}^H]$ *if the number of rounds goes to infinity.*[13]
- *For every $\theta_{ki} \in \Theta_{ki}$ and $\delta_{kj} \in \Delta_{kj}$, it holds that $f_i(\theta_{ki})$ and $p(\delta_{kj})$ are monotonically decreasing from round to round.*

The first property assures, that in the limit, there are no neighbored types that have a distance of more than an ϵ, which can be arbitrary small. That this is crucial to achieve

[12]Since we consider fully ordered type sets, 'between' is well-defined.

[13]The closure $\overline{S}$, of a subset S, consists of all points in S plus the limit points of S. The sets $\{\theta_{0i}^L, \dots, \theta_{0i}^H\}$ and $\{\delta_{0j}^L, \dots, \delta_{0j}^H\}$ contain all types that are added up to a considered round k.

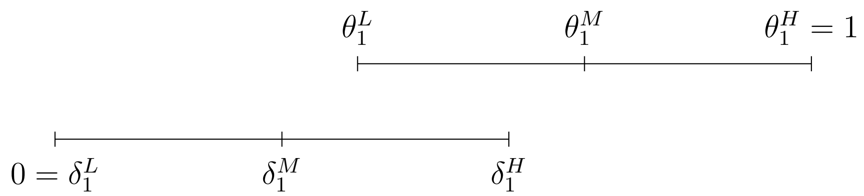
implementability in the limit can be seen in the section counterexamples later on.
We start in a situation with a binary type set, see Section 1.3, and call this setting 'round 0'. Successively, we add new types and index the new situations by round numbers $k = 1, 2, \ldots$. Exemplary we calculate the new efficient trade possibility condition for 'round 1', i.e., there are three types for the buyer and the seller. Hence, the buyer has the extended type set $\Theta_1 = \{\theta_1^L, \theta_1^M, \theta_1^H\}$, and the seller has the extended type set $\Delta_1 = \{\delta_1^L, \delta_1^M, \delta_1^H\}$.[14] Without loss of generality, we add the new types according to the following procedure and call it *uniform extension type I.*

Definition 1.3. (Uniform Extension Type I) *Consider given agent sets I_0 and J_0, for which every agent $i \in I_0$ has the same finite type set Θ_{0i} and for which every agent $j \in J_0$ has the same finite type set Δ_{0j}. The probabilities assigned to the possible types of any agent sum up to 1. This situation is called round 0. In every round $k = 1, 2, \ldots$ finitely many new types are added to each agents' type set. Thereby, new types θ_{ki} and δ_{kj} are positioned at the center between adjacent types from round $k = 0, \ldots, k-1$.*

In the following, we start with round 0, which is given by $I_0 = 1$ and $J_0 = 1$ and the type sets $\Theta_0 = \{\theta^L, \theta^H\}$ and $\Delta_0 = \{\delta^L, \delta^H\}$ and the respective probabilities, such that $f_0^L + f_0^H = 1$.[15]

Continuing, new types are introduced,in every 'round k', in the middle of the sub-intervals of type sets that existed already by 'round $k-1$'. With this procedure, the continuous type interval is approached, such that the first condition in Definition 1.2 is fulfilled. Since the probability of an existing type cannot increase, both conditions of Definition 1.2 are fulfilled. We consider the case where the introduced third type is positioned such that $\theta_1^M > \delta_1^H$.[16]

Figure 1.4: Three buyer and seller types



We can now derive which condition needs to be satisfied in order achieve efficient bilateral trade when the buyer and the seller have an extended type set with three types. Applying Proposition 1.4, the following Observation shows a condition for reaching efficient bilateral trade.

[14]Since $I = 1$ and $J = 1$, we drop the indices that label the agent for ease of notation.
[15]Remember that by Assumption 1.2, $f^L = p^H$ and $f^H = p^L$ and analogously for every round k.
[16]The case where $\theta_1^M < \delta_1^H$ can be found in Appendix 1.E.

Observation 1.6. *Without loss of generality, suppose parameters are as in Figure 1.4. For the 'Uniform Extension Type I' in 'round 1', there is a social choice function which is efficient, implementable as Bayes-Nash equilibrium, and yields non-negative material payoff for every type of player if and only if*

$$f_1^L > \frac{\delta_1^H - \theta_1^L}{\theta_1^M - \delta_1^M} \qquad \Leftrightarrow \qquad \frac{\frac{1}{2} f_1^L}{1 + \frac{1}{2} f_1^L} > d \; .$$

d still denotes $\frac{\delta^H - \theta^L}{\theta^H - \delta^L}$ and did not change from 'round 0' to 'round 1'. For the purpose of illustration, assume that every type has equal probability: In 'round 0' with a binary type set, the condition in Observation 1.1 states that $\frac{1}{2} > d$. After adding a third type, Observation 1.6 states that efficient bilateral trade is possible if $\frac{1}{7} > d$. So, even for only one additional type, the possibility condition for efficient trade gets more restrictive.

As for the binary type set, the minimal subsidy is important to evaluate whether efficient trade is possible. It is decisive, for which states of the economy trade takes place, i.e., for which combinations of buyer and seller types trade should take place, so that $q^*(\theta, \delta) = 1$. From Observation 1.6 we know that the high type buyer and the low type seller are not relevant. Intuitively, this is because even when a buyer with a lower valuation, i.e., θ_1^M, can trade with every possible type of seller, then this is as well true for high type buyers. By the same logic, the seller with the lowest valuation does not enter the implementability condition. Mathematically, for buyer's high type (and seller's low type), terms of the expected surplus effect and the expected information rent effect coincide and thus cancel out. Instead, the middle type (which is the lowest types that can trade with every possible seller) needs to be considered. When subsequent rounds are considered, we generalize this observation and analyze for which types the expected surplus effect and the expected information rent effect coincide. We define the buyer types and the seller types, for which the effects do not cancel out and thus are relevant for the evaluation of possibility.

When the type sets Θ_1 and Δ_1 are considered, expected surplus is given by

$$S(\cdot) = (\theta_1^M - \delta_1^M) f_1^M - d \; f_1^L (f_1^M + f_1^H) \; .$$

Expected information rents are given by

$$IR(\cdot) = (f_1^M + f_1^H)^2 (\theta_1^M - \delta_1^M + d) - f_1^H (\theta_1^M - \delta_1^M) \; .$$

Expected surplus is bigger than expected information rents if $\frac{\frac{1}{2} f_1^L}{1 + \frac{1}{2} f_1^L} > d$. This is the inequality we know from Observation 1.6.

The following Figure illustrates that implementation is more costly, i.e., the minimal subsidy goes up, if we add types to the setting. The lighter shaded area T_0, known from Figure 1.3, gives the set of tuples (f^L, d), for which trade with two types is possible. If we add a third type, the set T_1 (the darker shaded area) gives the combinations for which efficient trade is possible, where

$$T_1 = \left\{ (f^L, d) : f_1^L > \frac{\delta_1^H - \theta_1^L}{\theta_1^M - \delta_1^M} = \frac{2d}{1-d} \right\} ,$$

so that $T_1 \subseteq T_0$.

Given Definition 1.3, we can express the new type θ_1^M in terms of the types θ_1^L and θ_1^H. Therefore we can compare the sets T_0 and T_1 in a coordinate system, with axes labelled d and f^L.

Figure 1.5: Bilateral trade – Comparison between 'round 0' and 'round 1'

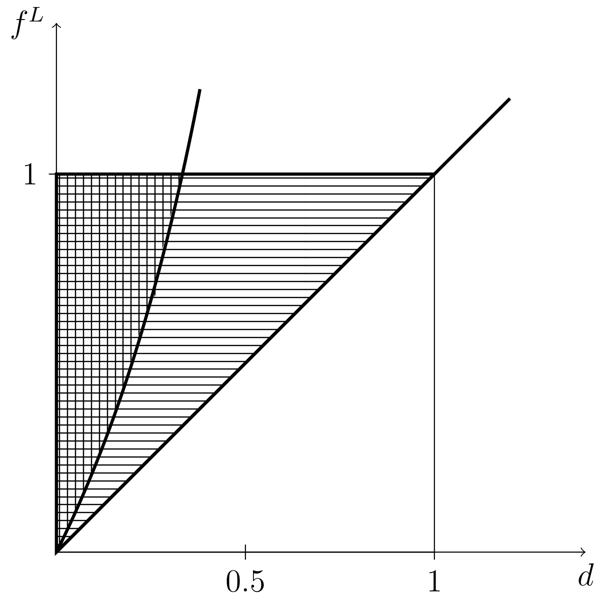


Figure 1.5 implies that if efficient trade is possible in round 1, then it is as well possible in 'round 0'. The opposite is, however, not true. Imagine that the ratio of overlapping is 0.5 in 'round 0'. Then, there is no value f^L can take, such that efficient bilateral trade is still possible.

Counterexamples. In order to illustrate the importance of Definition 1.2 and 1.3, we present examples that violate the 'Uniform Extension' procedure.

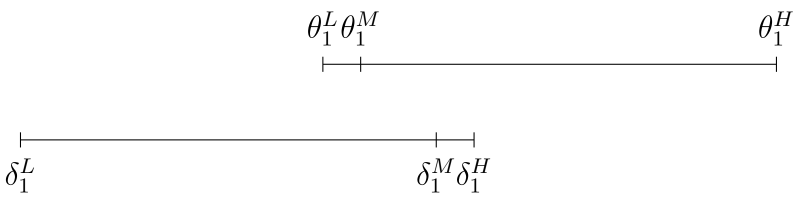
Observation 1.7. *Suppose parameters are as in Figure 1.4. Implementation in 'round 1' is less costly then in 'round 0' if*

$$f_1^L > \frac{2f_0^L}{1+f_0^L} \, .$$

For the purpose of illustration, assume that $f_0^L = \frac{1}{5}$. In 'round 0' with a binary type set, Observation 1.1 states that if $\frac{1}{5} > d$, then efficient bilateral trade is possible. Suppose now that $f_1^L = \frac{1}{2}$, according to Observation 1.7 efficient bilateral is possible if $\frac{1}{2} > d$. Hence, $T_1 \supseteq T_0$. If the probability of the low type increases over rounds, so that the second condition of Definition 1.2 is violated, then efficient bilateral trade gets less costly. For this result, the monotone hazard rate assumption plays no role. The probabilities on types f_1^H and f_1^M can be chosen such that the monotone hazard rate either holds or is violated.

Consider next the case, where the introduced third type is positioned such that $\theta_1^M < \delta_1^M$.

Figure 1.6: Violating the 'Uniform Extension' procedure



Applying Proposition 1.4, the following Observation shows a condition for reaching efficiency.

Observation 1.8. *Suppose parameters are as in Figure 1.6. There is a social choice function which is efficient, implementable as Bayes-Nash equilibrium, and yields non-negative material payoff for every type of agent if and only if*

$$f_1^L + f_1^M > d \; .$$

Comparing Observation 1.6 with Observation 1.8 illustrates that violating the first condition in Definition 1.2 makes it less costly to achieve efficient bilateral trade.

1.7.2 Introducing many types

To understand what leads to the Myerson and Satterthwaite (1983) impossibility result when the number of types is getting large, we must analyze the following characteristics of their model: Type sets are closed and connected sets.[17] Intuitively, this means that there are no gaps in the type set. When we add new types, we have to respect this characteristic of the model and achieve it in the limit.

[17] A set B is closed if it contains the limit points of every possible sequence $x_n \in B$, with $x_n \to x$, for $n \to \infty$. A set B is connected if it cannot be represented as the union of two or more disjoint non-empty open subsets.

Definition 1.4. (Uniform Extension Type II) *Consider the 'Uniform Extension Type I' Definition. Additionally, equal probability mass is put on all types, in every round.*

Making use of Definition 1.4, we can show that a discrete specification of the model approaches in the limit the continuous specification of the model.

Proposition 1.6. *Suppose Assumption 1.1 holds. Consider the 'Uniform Extension Type II' procedure. There is no social choice function which is efficient, implementable as Bayes-Nash equilibrium and yields non-negative material payoff for every type of agent if* $k \to \infty$.

The uniform extension procedure is one way to assure that the discrete specification of the bilateral trade problem approaches the continuous specification of Myerson and Satterthwaite (1983). In the proof we show that an increase in the number of types leads to an increase of the minimal subsidy. If sufficiently many types are introduced, then there exists a round $k < \infty$, where the minimal subsidy is positive, i.e., it is impossible to reach efficient bilateral trade.

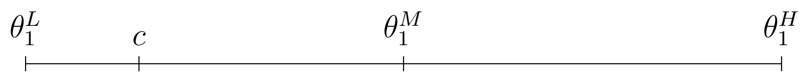
Public Good provision. Similar to the bilateral trade problem, the condition relates the type distribution to the relative position of the per capita costs for the public good. Therefore, we define

$$e = \frac{c - \theta^L}{\theta^H - c} .$$

Since $\theta^L < c < \theta^H$, the per capita costs c split the type set into two sub-intervals: the first, where the costs exceed the low valuation, and the second, where the costs lie below the high valuation. Thus, e is a measure for the net gains of free-riding. If consumer i has high preferences for the public good, he faces the following trade-off: If he announces his type truthfully, the public good will be provided for sure but he will have to pay higher transfers than if he is lying. However, if he understates his preferences, he will have to pay lower transfers but will risk that the public good will not be provided.

As for the bilateral trade setting, we use Definition 1.3 to show that implementation is getting more costly when more types are introduced. Analogously to the bilateral trade example, we consider the case where the third type is positioned such that $\frac{1}{2}\left(\theta_1^L + \theta_1^M\right) > c$.

Figure 1.7: Three consumer types



In order to efficiently provide the public good, when there are three consumer types, the inequality in the following Observation needs to be satisfied.

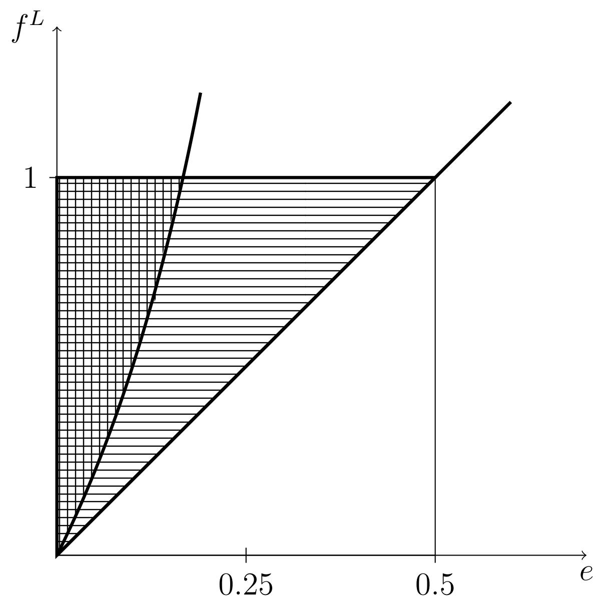
Observation 1.9. *Suppose parameters are as in Figure 1.7. There is a social choice function which is efficient, implementable as Bayes-Nash equilibrium and yields non-negative utility for every type of consumer, if and only if*

$$f_1^L > \frac{c - \theta_1^L}{\theta_1^M - c} .$$

As for the private good example, the expected surplus effect for the high valuation consumer equals the expected information rent effect. Thus, the parameter for the high-valuation buyer does not show up in Observation 1.9.

The comparison of Observations 1.4 and 1.9 shows that implementation is getting more costly with three consumer types. Even for one additional type, the possibility condition for efficient public good provision gets more restrictive. Graphically, a comparison of 'round 0' and 'round 1' resembles the graphic in the bilateral trade example.

Figure 1.8: Public good – Comparison between 'round 0' and 'round 1'



The gray shaded area D_0 in Figure 1.8 gives every combination of f^L and e, such that the public good is efficiently provided, i.e., Observation 1.9 holds,

$$D_0 = \left\{ (f^L, e) : f^L > e = \frac{c - \theta_0^L}{\theta_0^H - c} \right\} .$$

Analogously, D_1 gives the tuples (f^L, e), for which efficiency is achieved, if we have three possible types,

$$D_1 = \left\{ (f^L, e) : f^L > \frac{c - \theta_1^L}{\theta_1^M - c} = \frac{2e}{1-e} \right\} .$$

Hence, $D_1 \subseteq D_0$.

If we introduce a continuous type set to the public good setting with only two consumers, efficient public good provision is impossible.

Proposition 1.7. ***Börgers (2015):*** *If consumers' valuation for the public good are independently drawn from the interval $[\theta^L, \theta^H]$, then there is no Bayesian incentive compatible social choice function that is efficient and yields non-negative utility for every type of agent.*

We provide a proof that is an adaption of Myerson and Satterthwaite (1983) to the public good provision setting. Comparing Propositions 1.6 and 1.7, we can see that the assumption of a continuous type set makes both efficient bilateral trade as well as public good provision impossible. For the latter, the assumption of a large economy is hence not necessary for the impossibility to reach efficient public good provision.

1.7.3 General convergence

As we have seen in the previous subsections, there exist qualitatively different results concerning implementability in discrete and continuous environments. We analyze now how the results for discrete settings relate to the results in continuous settings. To answer this question, we reconstruct our given environment, such that we are able to link discrete and continuous environments technically and can specify how the respective implementation conditions are related. This reconstructed environment enables us to generalize our results concerning the effects of changing type sets or number of agents on implementability.

The economic environment in our paper contains a set of agents, a type set for every agent, and density functions. To combine the decisive factors for implementability, we bundle them in the definition of an *economy*. This generates a set of different economies whose elements can differ in the number of types, the number of agents, the density functions, or the allocation functions. On this set, we define sequences and the meaning of convergence. We frame requirements for this environment under which it is possible to align the implementability conditions in different economies. Thereby we find that under some assumptions concerning the *similarity* of these two environments (type sets, density functions and allocation functions), the implementability result in discrete settings approaches the result in the continuous setting. We are able to approach the Myerson and Satterthwaite (1983) result, even if we start with a parametrization, for which efficient trade is possible. We show conditions that need to be met in order to guarantee that we end up in an environment, that equals the setting of Myerson and Satterthwaite

(1983). This means, that it is not a sufficient condition to have infinitely many types in the bilateral trade setting to get the impossibility result. The position of the types in the type set and the probabilities matter. A first intuitive evidence for this relation is given by Observation 1.1. To show that even in case of infinitely many types efficient trade can take place, we will give some counterexamples to justify the conditions that we introduce afterwards. Thereafter we formulate the environment of economies for which we define the needed convergence and phrase the conditions, under which the infinite number of types is a sufficient condition for the impossibility of implementation. Even if the analysis of the counterexamples is done for the bilateral trade setting, the derived results apply to every independent private value setting. There will be a general result, when and under which conditions it is possible to link discrete and continuous (with respect to the type set) settings.

Consumption Economy.

Definition 1.5. *We define a consumption economy* (I, Θ, f, q) *as the tuple consisting of a consumer set* $I = \{1, \ldots, n\}$*, the Cartesian product of type sets (one type set for every consumer* i*)* $\Theta = \times_{i \in I} \Theta_i$ *with the corresponding density functions* $f_i : \Theta_i \to [0, 1]$*, where* $f := (f_1, \ldots, f_n)$*, and the allocation rule* $q : \Theta \to \mathbb{R}_+$*, which maps types into allocations.*

The economies can differ in the number of agents, the type set, the density function, or the allocation rule. To relate different economies, we have to introduce convergence to our set of economies. Especially, we want to link the finite and the infinite economy, where we consider finite and infinite type sets. For this, we define a sequence of consumption economies as given by $(I_k, \Theta_k, f_k, q_k)_{k \in \mathbb{N}}$, where I_k is a sequence of consumer sets, Θ_k a sequence of type set, f_k a sequence of density-functions, and q_k a sequence of allocation functions. We denote the infinite setting by $(I_\infty, \Theta_\infty, f_\infty, q_\infty)$, where we have an infinite number of agents and types. f_∞, q_∞ are the corresponding density or allocations functions. As our economy contains different formal objects, namely sets and functions, we have to define convergence component-wise.

Definition 1.6. *A sequence of consumption economies* $(I_k, \Theta_k, f_k, q_k)_{k \in \mathbb{N}}$ *converges to a limit economy* (I, Θ, f, q)*, if the sequences of consumers, type sets, density functions and allocation functions converge to the respective limit given by the limit consumption economy.*

When we look at the convergence of an economy, we require that every element of the economy converges. In the following we define in detail the components of an economy sequence and the corresponding concept of convergence.

Consumers. Let $I_k = \{1, \ldots, n_k\}$ be the set of consumers in round k. We say, that $(I_k)_{k \in \mathbb{N}} \to I_\infty$ if the number of consumers is increasing for $k \to \infty$.

I_k affects the dimension of the other components of an economy. Since we need a type set for every agent, the type set component in the economy consists of a Cartesian product over the type sets for every single agent. The domain of the density and the allocation function change accordingly.

Type sets. Fix a consumer set $I = \{1, \dots, n\}$. An element Θ_k of the sequence $(\Theta_k)_{k \in \mathbb{N}}$ is the Cartesian product over the consumers' type sets in round k. Since we assume symmetry, every consumer $i \in I$ in round k has the same type set. We call this type set Θ_{ki}. The component Θ_k of an economy in a sequence is hence given by $\Theta_k = (\Theta_{ki})^n$.

Since the type set component of the economy is a product over many type sets, we define the convergence of this Cartesian product element-wise, hence consumer-wise. This means, that for every consumer $i \in I$, the sequence of his type sets has to converge. We use the topological concept of density to define convergence.[18]

Definition 1.7. *Fix a consumer set I. If we write $(\Theta_k)_{k \in \mathbb{N}} \to \Theta$, this means $(\Theta_{ki})_{k \in \mathbb{N}} \to \Theta_i$, $k \to \infty$ for every $i \in I$. At that, Θ_i gives the limit of the sequence of type sets that is assigned to consumer $i \in I$. Thereby, it holds, that Θ_i is the limit of the sequence $(\Theta_{ki})_{k \in \mathbb{N}}$, if and only if, there exists a K, such that for all $k \geq K$, Θ_{ki} lies dense in $\Theta_i, \forall\, i \in I$.*

The sequence of the Cartesian products over type sets is generated by adding new types to every single type set of every consumer in a round k.[19] We add the same finite number of types to every one's type set. As mentioned above, whenever the number of consumers or the type sets change, the density function and the allocation rule has to change as well. In every round k and for every consumer $i \in I_k$, the sum over all probabilities of types is 1. Thus, we face the following issue: Whenever we add new types to type sets, we do not only need to assign a probability to them but also have to change probabilities of existing types, i.e., take probability mass of the old types and redistribute it to the new types. Thereby we make the assumption that for every single type, the probability cannot increase over rounds. This means, that the collected probabilities of the old types are shared exclusively among the newly added types.

Density Functions. Fix a consumer set $I = \{1, \dots, n\}$. The sequence $(f_k)_{k \in \mathbb{N}}$ consists of $f_k = (f_{k0}, \dots, f_{kn})$, where f_{ki} is the density function of consumer $i \in I$ in round k over his type set Θ_{ki}. The change in the type sets means, that we face a change in the domains of the density functions, such that the standard concept of convergence for function sequences cannot be used. Thus, for our purpose we require that for every type in the limit type set, that is also contained in one of the finite type sets, the difference between the probability assigned by the limit density function and the sequence of density functions has to become arbitrarily small if the number of rounds becomes large.

[18] A set $A \subseteq B$ lies dense in the set B, if $\overline{A} = B$, thereby $\overline{A} = A \cup \{\lim_n a_n, a_n \in A\}$.
[19] Mathematically this means: $\Theta_{ki} = \Theta_{(k-1)i} \cup \{\theta_m | m = 1, \dots, M; M < \infty\}$.

Definition 1.8. *Fix a consumer set $I = \{1, \ldots, n\}$. We say $(f_k)_{k\in\mathbb{N}} \to f, f = (f_1, \ldots, f_n) : \Theta \to [0, 1]^n$ if for any $\theta \in \Theta$, that lies also in Θ_k, for some $k < \infty$, there exists a $K < \infty$, such that for all $k \geq K$ it holds that $|f(\theta) - f_k(\theta)| < \epsilon$. This has to hold component-wise. Hence, for every consumer κ: $|f_i(\theta) - f_{ki}(\theta)| < \epsilon, i = 1, \ldots, n, k \geq K$.*

For the allocation functions, we do not have to change the outcomes in round k for the types that are contained in $\Theta_k \cap \Theta_{k-1}$. Since we consider first-best environments, we only have to define the outcomes assigned to new types. Since the domain changes for allocation functions, we require analogously for the density functions that for every type in the limit type set, that is also contained in one of the finite type sets, the difference between the outcome assigned by the limit allocation function and sequence of allocation functions has to become arbitrarily small if the number of rounds becomes large.

Allocation Functions. Fix a consumer set $I = \{1, \ldots, n\}$. The sequence of allocation functions is given by $(q_k)_{k\in\mathbb{N}} : (\Theta_k)_{k\in\mathbb{N}} \to (z_k)_{k\in\mathbb{N}}, z_k \in \mathbb{R}$, where $q_k : \Theta_k = (\Theta_{ki})^n \to \mathbb{R}_+$.

Definition 1.9. *We say $(q_k)_{k\in\mathbb{N}} \to q : \Theta \to \mathbb{R}_+$ if for any $\theta \in \Theta$, that lies also in Θ_k, for some $k < \infty$, there exists a $K < \infty$, such that for all $k \geq K$ it holds that $|q(\theta) - q_k(\theta)| < \epsilon$.*

For the public good example, we have already seen in Proposition 1.5 that a finite but sufficiently large number of consumers yields the same result as the Theorem by Mailath and Postlewaite (1990) for infinitely many consumers: The public good will not be provided. To apply these concepts to the private good example, we need the analogous definitions for the producer side. The respective definitions and concepts of convergence follow the same logic as for the consumer side and can be found in Appendix 1.D.

1.7.4 Convergence of type set

We are now able to bring together the implementation condition in discrete and continuous settings. The terms 'discrete' and 'continuous' refer to the set of types – for the private good application. We found that the implementability conditions in discrete environments approached the conditions in continuous environments such that the qualitative results concerning efficient implementation converge. Now, we can be more precise on what we mean by 'approach the results':

As we know from Section 1.5, efficient implementation is possible when consumers' expected transfers exceed firms' expected revenues. The expected transfers and revenues are calculated, taking the respective incentive compatibility and participation constraints into account. This logic also applies for continuous settings. To compare the implementability conditions for private goods in discrete and continuous environments, we increase the number of types, while holding the set of agents fixed. We find, that with the given definitions of economies and convergence, the expected payments for a finite type set converge to the expected payments in the continuous economy. The payments that we get for a mechanism (q, t) in a continuous environment that fulfills incentive compatibility and participation constraints can be found in equation (1.10). With the same arguments, we get the convergence of the revenues for the producer side, see Appendix 1.D.

Proposition 1.8. *Let $(I, \Theta_k, f_k, q_k)_{k\in\mathbb{N}} \to (I, \Theta_\infty, f_\infty, q_\infty)$, for $k \to \infty$. Then it holds that for every consumer $i \in I$: For every $\epsilon > 0 \quad \exists K : \forall\, k \geq K$*

$$\Bigg|\mathbb{E}_{(\theta_{ki})}\left[\left(v(\theta_{ki}, q_{ki}(\theta,\delta)) - \{v(\theta_{ki}^{+}, q_{ki}(\theta,\delta)) - v(\theta_{ki}, q_{ki}(\theta,\delta))\}\frac{1-F_k(\theta_{ki})}{f_k(\theta_{ki})}\right)\right] - \mathbb{E}_{(\theta_{\infty i})}\left[v(\theta_{\infty i}, q_{\infty i}(\theta,\delta)) - v_1(\theta_{\infty i}, q_{\infty i}(\theta))\frac{1-F_\infty(\theta_{\infty i})}{f_\infty(\theta_{\infty i})}d\theta_{\infty i}\right]\Bigg| < \epsilon.$$

Thus, for the private good application, we get equivalent conditions for efficient implementation of finite but sufficiently large type sets and the infinite type set.

For public goods, we know from Mailath and Postlewaite (1990) that for an infinite number of consumer and a finite type set, the public good will never be provided if incentive compatibility, participation, and the resource constraints have to be fulfilled. As shown in Proposition 1.6, we find that for a finite but sufficiently large set of consumers, the same result is true: The public good will not be provided. While increasing the number of agents, the number of types remains unchanged.

Summarizing, we changed the element of the economies in the private and the public good example that had infinite dimension. For the bilateral trade application this is given by the type set, whereas it is the set of agents in the public good example. If we approach these infinite components of the economies (and the resulting changes in the density and allocation functions) by a sequence of finite but increasing elements, we can generate the same implementation result for the finite environments as for the original infinite cases.

The stated convergence result does also apply to every possible application of the independent private values model that fulfills the conditions. Thus, this result enables us to link results for discrete and infinite settings. Thereby, either type sets or agent sets (or both) can be varied from a finite number of elements to an infinite set.

1.8 Concluding remarks

In this paper, we investigate the independent private values model when types are discrete. We use this model for the analysis of how impossibility results by Myerson and Satterthwaite (1983) and Mailath and Postlewaite (1990) are affected by the specification of the number of individuals and the specification of the type set. The existing literature on this topic neglects the question of how the discrete specification and the continuous specification of the independent private values model relate to each other. Our analysis provides a framework to study the convergence.

Our analysis yields the following key insights: First, the impossibility results for efficient bilateral trade and efficient public good provision vanish when a binary type set is considered. Moreover, we find that the Mailath and Postlewaite (1990) result extends to any model with a discrete set of types. The Myerson and Satterthwaite (1983) result, by contrast, extends only to a model with a large but finite number of types.

Second, the discrete version of the independent private values model leads to the same outcomes as the continuous version of the model if many types are introduced in the right way. We discuss various factors that have an influence on the convergence. This analysis does not support the presumption that the increase in the type set alone leads to the convergence of both model specifications.

The analysis was made possible by a combination of insights from the non-linear pricing literature and mechanism design. We believe that the applicability of a large class of problems that have been studied in the empirical economics and behavioral economics literature can be simplified by using the discrete specification of the independent private values model.

Appendix 1.A Preliminaries

Lemma 1.1. *For all i, the incentive constraints in (IC_C) hold if the following local incentive constraints are satisfied: For any $l < s$,*

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \,, \tag{1.11}$$

and for all $l > 0$,

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^{l-1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l-1})) - T(\theta_i^{l-1}) \,. \tag{1.12}$$

Moreover, the local incentive constraints imply that, for all i, $q(\theta_{-i}, \theta_i^l) \geq q_i(\theta_{-i}, \theta_i^{l-1})$, for all $l > 1$.

Proof: We first show that $q_i(\theta_{-i}, \theta_i^{l+1}) \geq q_i(\theta_{-i}, \theta_i^l)$ for each i and each l. Equation (1.12) as stated in the Lemma for $\theta_i = \theta_i^{l+1}$:

$$V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \geq V(\theta_i^l \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \,.$$

Adding equation (1.11) as stated above yields:

$$\begin{aligned} V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) + V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^l) - T(\theta_i^{l+1}) \geq \\ V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) + V(\theta_i^l \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^{l+1}) - T(\theta_i^l) \end{aligned}$$

$$\begin{aligned} \Leftrightarrow V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) + V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) \geq \\ V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) + V(\theta_i^l \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^l)) \end{aligned}$$

$$\begin{aligned} \Leftrightarrow V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^l \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^l)) \geq \\ V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) \end{aligned}$$

Since $\theta_i' \geq \theta_i$ implies that $v_2(\theta_i', q_i) \geq v_2(\theta_i, q_i)$, this holds, as long as $q_i(\theta_{-i}, \theta_i^{l+1}) \geq q(\theta_{-i}, \theta_i^l)$.
We show that equation (1.11) implies that

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^{l+2} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+2})) - T(\theta_i^{l+2}) \,.$$

To see this, rewrite equation (1.11) as

$$
\begin{aligned}
V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq \\
& V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \\
& - \left[V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1}))\right] .
\end{aligned}
$$

Since $q_i(\theta_{-i}, \theta_i^{l+2}) \geq q_i(\theta_{-i}, \theta_i^{l+1})$, we have

$$
\begin{aligned}
V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq \\
& V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \\
& - \left[V(\theta_i^{l+2} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+2})) - V(\theta_i^{l+2} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+2}))\right] .
\end{aligned}
$$

Moreover, condition (1.11) for $\theta_i = \theta_i^{l+1}$ is

$$
V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \geq V(\theta_i^{l+2} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+2})) - T(\theta_i^{l+1}) .
$$

Adding the last two inequalities yields

$$
V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^{l+2} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+2})) - T(\theta_i^{l+1}) .
$$

Hence, an individual with preference parameter θ_i^l does not benefit from announcing θ_i^{l+2}. Iterating this argument once more establishes that this individual does neither benefit from announcing θ_i^{l+3}, etc. The proof that an individual with preference parameter θ_i^l does not benefit from announcing θ_i^{l+j}, for any $j \geq 1$ is analogous and left to the reader.

Lemma 1.2. *Suppose that, for some individual i, all local downward incentive compatibility constraints are binding and that $q_i(\theta_{-i}, \theta_i^l) \geq q_i(\theta_{-i}, \theta_i^{l-1})$, for all $l > 1$. Then all incentive constraints of i are satisfied.*

Proof: If all local downward incentive constraints are binding for individual i, this implies that, for all $l \geq 1$,

$$
T(\theta_i^l) = \sum_{k=1}^{l} \{V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^{k-1} \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^{k-1}))\} + T(\theta_i^0) .
$$

Using that $v_i(\theta_i^0, q_i) = 0$, for all $l > 0$, the equation can be equivalently written as

$$
\begin{aligned}
T(\theta_i^l) = & V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) \\
& - \sum_{k=0}^{l-1} \{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} + T(\theta_i^0) .
\end{aligned}
$$

To establish incentive compatibility, Lemma 1.1 implies that it suffices to show that all local upward incentive constraints are satisfied, i.e., for all l,

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \ ,$$

or equivalently,

$$\begin{aligned} V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq & \\ & V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - T(\theta_i^{l+1}) \\ & - \left[V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1}))\right] \ . \end{aligned}$$

By $T(\theta_i^l) = V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - \sum_{k=0}^{l-1}\{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} + T_i(\theta_i^0)$, this inequality can be written as

$$\begin{aligned} \sum_{k=0}^{l-1}\{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} \geq & \\ & \sum_{k=0}^{l}\{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} \\ & - \left[V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1}))\right] \ , \end{aligned}$$

or

$$\begin{aligned} V(\theta_i^{l+1} \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^{l+1})) - V(\theta_i^{l+1} \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^{l+1})) \geq & \\ V(\theta_i^l \mid \theta_i^{l+1}, q_i(\theta_{-i}, \theta_i^l)) - V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)). & \end{aligned}$$

As $\theta_i^{l+1} > \theta_i^l$, the inequality is satisfied if

$$q_i(\theta_{-i}, \theta_i^{l+1}) \geq q_i(\theta_{-i}, \theta_i^l) \ .$$

These monotonicity constraints are satisfied by assumption.

Lemma 1.3. *If for individual i, all local downward incentive compatibility constraints are binding, then the expected utility of individual i from ex ante perspective is given by*

$$\mathbb{E}_{(\theta)}[v(\theta_i, q_i(\theta)) - t_i(\theta)] = \mathbb{E}_{(\theta)}\left[\{v(\theta_i^+, q_i(\theta)) - v(\theta_i, q_i(\theta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right] - T(\theta_i^0) \ .$$

Proof: Equation

$$\begin{aligned} T(\theta_i^l) = & V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) \\ & - \sum_{k=0}^{l-1}\{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} + T(\theta_i^0) \ , \end{aligned}$$

in the proof of Lemma 1.2 and the law of iterated expectations imply that,

$$\begin{aligned}
\mathbb{E}_{(\theta)}[t_i(\theta)] &= \sum_{j=0}^{s} f_i^j T(\theta_i^j) \\
&= \sum_{j=0}^{s} f_i^j \mathbb{E}_{(\theta_{-i})}[v(\theta_i^j, q_i(\theta_{-i}, \theta_i^j))] \\
&\quad - \sum_{j=1}^{s} f_i^j \sum_{k=0}^{j-1} \{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} + T(\theta_i^0) \\
&= \mathbb{E}_{(\theta)}[v(\theta_i, q_i(\theta))] + T(\theta_i^0) \\
&\quad - \sum_{j=1}^{s} f_i^j \sum_{k=0}^{j-1} \{V(\theta_i^k \mid \theta_i^{k+1}, q_i(\theta_{-i}, \theta_i^k)) - V(\theta_i^k \mid \theta_i^k, q_i(\theta_{-i}, \theta_i^k))\} \\
&= \mathbb{E}_{(\theta)}[v(\theta_i, q_i(\theta))] + T(\theta_i^0) \\
&\quad - \sum_{j=1}^{s} \left(1 - \sum_{k=0}^{j} f_i^j\right) \{V(\theta_i^j \mid \theta_i^{j+1}, q_i(\theta_{-i}, \theta_i^j)) - V(\theta_i^j \mid \theta_i^j, q_i(\theta_{-i}, \theta_i^j))\} \\
&= \mathbb{E}_{(\theta)}[v(\theta_i, q_i(\theta))] + T(\theta_i^0) \\
&\quad - \left[\sum_{j=1}^{s} f_i^k \{V(\theta_i^j \mid \theta_i^{j+1}, q_i(\theta_{-i}, \theta_i^j)) - V(\theta_i^j \mid \theta_i^j, q_i(\theta_{-i}, \theta_i^j))\} \frac{1 - \sum_{k=0}^{j} f_i^k}{f_i^j}\right] \\
&= \mathbb{E}_{(\theta)}[v(\theta_i, q_i(\theta))] - \mathbb{E}_{(\theta)}\left[\{v(\theta_i^+, q_i(\theta)) - v(\theta_i, q_i(\theta))\} \frac{1 - F(\theta_i)}{f(\theta_i)}\right] + T(\theta_i^0)
\end{aligned}$$

Lemma 1.4. *For all i, if the (PC_C) is satisfied for $\theta_i = \theta_i^0$, then it is satisfied as well for all $\theta_i \neq \theta_i^0$.*

Proof: Let $\theta_i \neq \theta_i^0$. Then, by the incentive constraints in (1.12)

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - T(\theta_i^l) \geq V(\theta_i^0 \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^0)) - T(\theta_i^0) \; .$$

Moreover, $\theta_i > \theta_i^0$ implies that the right-hand side of this inequality exceeds

$$V(\theta_i^0 \mid \theta_i^0, q_i(\theta_{-i}, \theta_i^0)) - T(\theta_i^0) \; ,$$

which is non-negative by (PC_C) for $\theta_i = \theta_i^0$. This proves that (PC_C) is not binding for $\theta_i \neq \theta_i^0$.

Lemma 1.5. *Let q be an arbitrary given provision rule. Consider the problem of choosing* *a mechanism* $(t_1, ..., t_n)$ *in order to maximize total transfers*

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n} t_i(\theta)\right] \ ,$$

subject to the incentive compatibility constraints in (1.12) and the interim participation constraints in (PC_C)*. At a solution to this problem, the participation constraint in (**??**) is binding for* $\theta_i = \theta_i^0$ *and is slack otherwise.*

Proof: By Lemma 1.4 we only need to show that it is binding for $\theta_i = \theta_i^0$. We show that it is possible to increase the expected payments of individual i in an incentive compatible way if, for some i, the participation constraint for $\theta_i = \theta_i^0$ does not hold as an equality. It is instructive to rewrite the incentive compatibility constraints in (1.12) as follows: For each i, for each $\theta_i^l \in \Theta_i$, and for each $\hat{\theta}_i \neq \theta_i^l \in \Theta_i$,

$$V(\theta_i^l \mid \theta_i^l, q_i(\theta_{-i}, \theta_i^l)) - V(\hat{\theta}_i^l \mid \theta_i^l, q_i(\theta_{-i}, \hat{\theta}_i^l)) \geq T(\theta_i^l) - T(\hat{\theta}_i) \ .$$

Consider a new payment rule for individual i such that for each $\theta_i \in \Theta_i$, $T(\theta_i^l)$ increases by some $\epsilon > 0$, this implies that the right-hand side of the incentive constraints states above remains constant, i.e., the increase of i's expected payments does not violate the incentive compatibility. Since revenue increases in the expected payments of individual i, the revenue maximizing mechanism must be such that a biding participation constraint for $\theta_i = \theta_i^0$ prevents a further increase of individual i's payments.

Lemma 1.6. *Let* q *be an arbitrary given provision rule. Consider the "relaxed problem" of choosing a mechanism* $(t_1, ..., t_n)$ *in order to maximize total transfers*

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n} t_i(\theta)\right] \ ,$$

subject to the downward incentive compatibility constraints in (1.12) and the ex interim participation constraints in (PC_C)*. At a solution to this problem, all downward incentive constraints are binding, and the participation constraint in (**??**) is binding for* $\theta_i = \theta_i^0$ *and slack otherwise.*

Proof: It is straightforward to verify that, for all i, all downward incentive constraints are binding. Otherwise the expected payments of some individual could be increased without violating any one of the constraints of the relaxed problem. It remains to be shown that, for all i, the participation constraint in (PC_C) is binding for $\theta_i = \theta_i^0$ and is slack otherwise. By Lemma 1.4 we only need to show that, for all i, the participation constraint in (PC_C) is binding for $\theta_i = \theta_i^0$. Suppose otherwise, then is was possible to increase $T(\theta_i^0)$ without violating any constraint.

Lemma 1.7. *Let q be a given provision rule with the property that for all i, and all l, the monotonicity constraints $q_i(\theta_{-i}, \theta_i^l) > q_i(\theta_{-i}, \theta_i^{l-1})$ are satisfied. Consider the problem of choosing $(t_1, ..., t_n)$ in order to maximize the total transfers*

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n} t_i(\theta)\right] ,$$

subject to the incentive compatibility constraints in (1.12) and the interim participation constraints in (??). The maximal transfers at a solution to this problem is equal to

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n}\left[v(\theta_i, q_i(\theta)) - \{v(\theta_i^+, q_i(\theta)) - v(\theta_i, q_i(\theta))\}\frac{1 - F(\theta_i)}{f(\theta_i)}\right]\right] .$$

Proof: First, consider the "relaxed problem" of maximizing expected transfers subject to the local downward incentive constraints in (1.12) and the participation constraints for $\theta_i = \theta_i^0$. The arguments in the proofs of Lemma 1.4 – 1.6 imply that, for all i, all local downward incentive constraints as well as the ex interim participation constraints are binding for $\theta_i = \theta_i^0$.

Since the given provision rule q satisfies the monotonicity constraints in $q_i(\theta_{-i}, \theta_i^l) > q_i(\theta_{-i}, \theta_i^{l-1})$ for all i and all l, Lemma 1.2 implies that all incentive compatibility constraints are satisfied at a solution to the relaxed problem. Hence, the solution to the relaxed problem is the revenue maximizing mechanism.

Given that all local downward incentive compatibility constraints are binding, Lemma 1.3 implies that, for all i,

$$\mathbb{E}_{(\theta)}\left[t_i(\theta)\right] = \mathbb{E}_{(\theta)}\left[v(\theta_i, q_i(\theta)) - \{v(\theta_i^+, q_i(\theta)) - v(\theta_i, q_i(\theta))\}\frac{1 - F(\theta_i)}{f(\theta_i)}\right] + T_i(\theta_i) .$$

Since the participation constraints are binding, for all i, whenever $\theta_i = \theta_i^0$, we have $T_i(\theta_i^0) = -\underline{u}_i$, for all i, and hence

$$\mathbb{E}_{(\theta)}\left[\sum_{i=1}^{n} v(\theta_i, q_i(\theta, \delta)) - \sum_{i=1}^{n}\{v(\theta_i^+, q_i(\theta)) - v(\theta_i, q_i(\theta))\}\frac{1 - F(\theta_i)}{f(\theta_i)}\right] - \sum_{i=1}^{n}\underline{u}_i .$$

Lemma 1.8. *For all j, the incentive constraints in (IC_F) hold if the following local incentive constraints are satisfied: For any $l < r$,*

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^{l+1}) - K(\delta_j^{l+1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l+1})) , \quad (1.13)$$

and for all $l > 1$,

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1})) \;. \tag{1.14}$$

Moreover, the local incentive constraints imply that, for all j, $y_j(\delta_{-j}, \delta_j^l) \leq y_j(\delta_{-j}, \delta_j^{l-1})$, for all $l > 2$.

Proof: We first show that $y_j(\delta_{-j}, \delta_j^{l+1}) \leq y_j(\delta_{-j}, \delta_j^l)$ for each j and each l. Equation (1.13) as stated in the Lemma for $\delta_j = \delta_j^{l-1}$:

$$R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) \geq R(\delta_j^l) - K(\delta_j^l \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^l)) \;.$$

Adding equation (1.14) as stated above yields:

$$\begin{aligned} K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \\ K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^l \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^l)). \end{aligned}$$

Since $\delta_l' > \delta_l$, this implies htat $k_2(\delta_j', y_j) \geq k_2(\delta^l, y_l)$, as long as $y_j(\delta_{-j}, \delta_j^l) \leq y_j(\delta_{-j}, \delta_j^{l-1})$. We show that equation (1.14) implies that

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^{l-2}) - K(\delta_j^{l-2} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-2})) \;.$$

To see this rewrite equation (1.14) as

$$\begin{aligned} R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \\ & R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) \\ & + [K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1}))] \;. \end{aligned}$$

Since $y_j(\delta_{-j}, \delta_j^{l-2}) \geq y_j(\delta_{-j}, \delta_j^{l-1})$, we have

$$\begin{aligned} R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \\ & R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) \\ & + K(\delta_j^{l-2} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-2})) - K(\delta_j^{l-2} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-2})) \;. \end{aligned}$$

Moreover, condition (1.14) for $\delta_j = \delta_j^{l-1}$ is

$$R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) \geq R(\delta_j^{l-2}) - K(\delta_j^{l-2} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-2})) \;.$$

Adding the two last inequalities yields

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^{l-2}) - K(\delta_j^{l-2} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-2})) \;.$$

Hence, a firm with technology parameter δ_j^l does not profit from announcing δ_j^{l-2}. Iterating this argument once more establishes that this firm does neither profit from announcing δ_j^{l-3}, etc. The proof that a firm with technology parameter δ_j^l does not profit from announcing δ_j^{l-j}, for any $j \geq 1$ is analogous and left to the reader.

Lemma 1.9. *Suppose that, for some firm j, all local upward incentive constraints are binding and that $y_j(\delta_{-j}, \delta_j^l) \geq y_j(\delta_{-j}, \delta_j^{l-1})$, for all $l > 2$. Then all incentive compatibility constraints are satisfied.*

Proof: If all local upward incentive constraints are binding for firm j, this implies that, for all $l < r$,

$$R(\delta_j^l) = \sum_{k=l}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^{k+1} \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^{k+1}))\} + R(\delta_j^r) \ .$$

For all $l > 2$, the equation can equivalently be written as

$$\begin{aligned} R(\delta_j^l) =& K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \\ &+ \sum_{k=l+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} + R(\delta_j^r) \ . \end{aligned}$$

To establish incentive compatibility, Lemma 1.8 implies that it suffices to show that all local downward incentive compatibility constraints are satisfied, i.e., for all l,

$$R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1})) \ ,$$

or equivalently,

$$\begin{aligned} R(\delta_j^l) - & K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \\ & R(\delta_j^{l-1}) - K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) \\ & + \left[K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1}))\right] \ . \end{aligned}$$

By equation

$$\begin{aligned} R(\delta_j^l) =& K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \\ &+ \sum_{k=l+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} + R(\delta^r) \ , \end{aligned}$$

this inequality can be written as:

$$\begin{aligned}\sum_{k=l+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} \geq \\ \sum_{k=l}^{r-2}\{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} \\ + \left[K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1}))\right] ,\end{aligned}$$

or

$$\begin{aligned}K(\delta_j^l \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^l)) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq \\ K(\delta_j^{l-1} \mid \delta_j^{l-1}, y_j(\delta_{-j}, \delta_j^{l-1})) - K(\delta_j^{l-1} \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^{l-1})) \ .\end{aligned}$$

This inequality is fulfilled if

$$y_j(\delta_{-j}, \delta_j^{l-1}) \geq y_j(\delta_{-j}, \delta_j^l) \ .$$

These monotonicity constraints are fulfilled by assumption.

Lemma 1.10. *If for firm j, all local upward incentive constraints are biding, then the expected profit of firm j from ex ante perspective is given by*

$$\mathbb{E}_{(\delta)}[r_j(\delta) - k(\delta_j, y_j(\delta))] = \mathbb{E}_{(\delta)}\left[\{k(\delta_j, y_j(\delta)) - k(\delta_j^-, y_j(\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right] + R(\delta_j^r) \ .$$

Proof: Equation

$$\begin{aligned}R(\delta_j^l) = & K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \\ & + \sum_{k=l+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} + R(\delta^r) \ ,\end{aligned}$$

in the proof of Lemma 1.9 and the law of iterated expectation imply that

$$
\begin{aligned}
\mathbb{E}_{(\delta)}[r_j(\delta)] &= \sum_{i=1}^{r} p_j^i R(\delta_j^i) \\
&= \sum_{i=1}^{r} p_j^i \Big[\mathbb{E}_{(\delta_{-j})}[k(\delta_j^i, y_j(\delta_{-j}, \delta_j^i))] \\
&\quad + \sum_{k=i+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} - R(\delta_j^r)\Big] \\
&= \quad \mathbb{E}_{(\delta)}[k(\delta_j, y_j(\delta))] - R(\delta_j^r) \\
&\quad \sum_{i=1}^{r} p_j^i \sum_{k=i+1}^{r-1} \{K(\delta_j^k \mid \delta_j^k, y_j(\delta_{-j}, \delta_j^k)) - K(\delta_j^k \mid \delta_j^{k-1}, y_j(\delta_{-j}, \delta_j^k))\} \\
&= \quad \mathbb{E}_{(\delta)}[k(\delta_j, y_j(\delta))] - R(\delta_j^r) \\
&\quad + \sum_{i=1}^{r} \sum_{k=i}^{r-1} p_j^{i-1} \{K(\delta_j^i \mid \delta_j^i, y_j(\delta_{-j}, \delta_j^i)) - K(\delta_j^i \mid \delta_j^{i-1}, y_j(\delta_{-j}, \delta_j^i))\} \\
&= \quad \mathbb{E}_{(\delta)}[k(\delta_j, y_j(\delta))] + \mathbb{E}_{(\delta)}\left[\{k(\delta_j, y_j(\delta)) - k(\delta_j^-, y_j(\theta, \delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right] - R(\delta_j^r)
\end{aligned}
$$

Lemma 1.11. *For all j, if the (PC_F) is satisfied for $\delta_j = \delta_j^r$, then it is satisfied as well for $\delta_j \neq \delta_j^r$.*

Proof: Let $\delta_j \neq \delta_j^r$. Then by (1.13)

$$
R(\delta_j^l) - K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) \geq R(\delta_j^r) - K(\delta_j^r \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^r)) \; .
$$

Moreover, $\delta_j < \delta_j^r$ implies that the right-hand side of this inequality exceeds

$$
R(\delta_j^r) - K(\delta_j^r \mid \delta_j^r, y_j(\delta_{-j}, \delta_j^r)) \; ,
$$

which is non-negative by (PC_F) for $\delta_j = \delta_j^r$. This proves that (PC_F) is not binding for all $\delta_j \neq \delta_j^r$.

Lemma 1.12. *Let y be an arbitrary production rule. Consider the problem of choosing a mechanism $(r_1, ..., r_m)$ in order to minimize revenue*

$$
\mathbb{E}_{(\delta)}\left[\sum_{j=1}^{m} r(\delta)\right] \; ,
$$

subject to the incentive compatibility constraints in (1.13) and the interim participation constraints (PC_F). At a solution to this problem, the participation constraint in (PC_F) is binding for $\delta_j = \delta_j^r$ and slack otherwise.

Proof: By Lemma 1.11 we only need to show that it is binding for $\delta_j = \delta_j^r$. We show that it is possible to decrease expected revenues of firm j in an incentive compatible way if, for some j, the participation constraints for $\delta_j = \delta_j^r$ does not hold as an equality. It is instructive to rewrite the incentive compatibility constraints in (IC_F) as follows: For each j, for each $\delta_j^l \in \Delta_j$, and for each $\hat{\delta}_j \neq \delta_j^l \in \Delta_j$,

$$R(\delta_j^l) - R(\hat{\delta}_j) \geq K(\delta_j^l \mid \delta_j^l, y_j(\delta_{-j}, \delta_j^l)) - K(\hat{\delta}_j \mid \delta_j^l, y_j(\delta_{-j}, \hat{\delta}_j)) \ .$$

Consider a new revenue rule for firm j such that for every $\delta_j \in \Delta_j$, $R(\delta_j^l)$ decreases by some $\epsilon > 0$. This implies that the left-hand side of the incentive constraint remains constant, i.e., the decrease of j's expected revenue does not violate the incentive compatibility. Since revenue increase in the expected transfers to firm j, the revenue minimizing mechanism must be such that a binding participation constraint for $\delta_j = \delta_j^r$ prevents a further increase of firms j's revenues.

Lemma 1.13. *Let y be an arbitrary given provision rule. Consider the "relaxed problem" of choosing a mechanism $(r_1, ..., r_m)$ in order to minimize total revenue*

$$\mathbb{E}_{(\delta)}\left[\sum_{j=1}^{m} r_j(\delta)\right] \ ,$$

subject to the upward incentive compatibility constraints in (IC_F) and the interim participation constraints (PC_F). At a solution to this problem, all upward incentive constraints are biding, and the participation constraint in (PC_F) is binding for $\delta_j = \delta_j^r$ and slack otherwise.

Proof: It is straightforward to verify that, for all j, all upward incentive constraints are binding. Otherwise, the expected revenue of some firm j could be decreased without violating any one of the constraints of the relaxed problem. It remains to be shown that, for all j, the participation constraint in (PC_F) is binding for $\delta_j = \delta_j^r$. Suppose otherwise, then it was possible to decrease $R(\delta_j^r)$ without violating any constraint.

Lemma 1.14. *Let y be a given provision rule with the property that for all j, and all l, the monotonicity constraints $y_j(\delta_{-j}, \delta_j^l) \leq y_j(\delta_{-j}, \delta_j^{l-1})$ are satisfied. Consider the problem of choosing $(r_1, ...r_m)$ in order to minimize the total revenue*

$$\mathbb{E}_{(\delta)}\left[\sum_{j=1}^{m} r_j(\delta)\right] \ ,$$

subject to the incentive compatibility constraints in (IC_C) and the interim participation *constraints (PC_C). The minimal revenue at a solution to this problem is equal to*

$$\mathbb{E}_{(\delta)}\left[\sum_{j=1}^{m}\{k(\delta_j, y_j(\delta)) - k(\delta_j^-, y_j(\delta))\}\frac{P(\delta_j)}{p(\delta_j)} - \sum_{j=1}^{m} k(\delta_j, y_j(\delta))\right] - \sum_{j=1}^{m} R(\delta^r)\ ,$$

Proof: First, consider the "relaxed problem" of minimizing expected revenues subject to the local incentive constraints in (1.13) and the participation constraints for $\delta_j = \delta_j^r$. The arguments in the proofs of Lemma 1.11 – 1.13 imply that, for all j, all local upward incentive constraints as well as the participation constraint for $\delta_j = \delta_j^r$ are binding.

Since the given production rule y satisfies the monotonicity constraints in $y_j(\delta_{-j}, \delta_j^l) \leq y_j(\delta_{-j}, \delta_j^-)$ for all j and all l, Lemma 1.9 implies that all incentive compatibility constraints are satisfied at a solution to the relaxed problem. Hence the solution to the relaxed problem is the revenue minimizing mechanism.

Given that all local upward incentive compatibility constraints are binding, Lemma 1.10 implies that, for all j,

$$\mathbb{E}_{(\delta)}\left[r_j(\delta)\right] = \mathbb{E}_{(\delta)}\left[\{k(\delta_j, y_j(\delta)) - k(\delta_j^{l-1}, y_j(\delta))\}\frac{P(\delta_j)}{p(\delta_j)} - k(\delta_j, y_j(\delta))\right] - R(\delta_j^r)\ .$$

Since the participation constraints are binding, for all j, whenever $\delta_j = \delta_j^r$, we have that $r(\delta_j^r) = \underline{\pi}_j$, for all j, and hence

$$\mathbb{E}_{(\delta)}\left[\sum_{j=1}^{m}\{k(\delta_j, y_j(\delta)) - k(\delta_j^-, y_j(\delta))\}\frac{P(\delta_j)}{p(\delta_j)} - \sum_{j=1}^{m} k(\delta_j, y_j(\delta))\right] - \sum_{j=1}^{m} \underline{\pi}_j\ .$$

Appendix 1.B Proofs of Propositions and Corollaries

Proof of Proposition 1.3.
Consider the 'first relaxed problem' of maximizing expected transfers subject to local downward incentive compatibility constraints in (1.12) and interim participation constraints in (PC_C). The arguments in the proofs of Lemma 1.4, 1.5 and 1.6 imply that, for all i, all local downward incentive compatibility constraints as well as the interim participation constraints are binding for $\theta_i = \theta^0$. In these Lemmata the provision rule for the publicly provided good is not taken as given. However, this does not affect the logic of the argument.
Given that all local incentive compatibility constraints are biding, Lemma 1.2 implies that, for all i,

$$\begin{aligned}&\mathbb{E}_{(\theta,\delta)}[t_i(\theta,\delta)]\\ =&\mathbb{E}_{(\theta,\delta)}\left[\left(v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] + T(\theta_i^0)\ .\end{aligned}$$

Since the participation constraints are binding, for all i, whenever $\theta_i = \theta_i^0$, we have $T(\theta_i^0) = -\underline{u}_i$, for all i, and hence

$$\begin{aligned}&\mathbb{E}_{(\theta,\delta)}[t_i(\theta,\delta)]\\ =&\mathbb{E}_{(\theta,\delta)}\left[\left(v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] - \underline{u}_i\ .\end{aligned}$$

Second, consider the 'second relaxed problem' of minimizing revenues subject to the local downward incentive compatibility constraints in (1.14) and the interim participation constraints in (PC_F). The arguments in the proofs of Lemmata 1.9, 1.11 and 1.12 imply that, for all j, all local upward incentive compatibility constraints as well as the interim participation constraints are binding for $\delta_j = \delta^r$. In these Lemmata the production rule for the good is not taken as given. However, this does not affect the logic of the argument.
Given all local incentive compatibility constraints are binding, Lemma 1.9 implies, for all j,

$$\begin{aligned}&\mathbb{E}_{(\theta,\delta)}[r_j(\theta,\delta)]\\ =&\mathbb{E}_{(\theta,\delta)}\left[\left(k(\delta_j,y_j(\theta,\delta)) + \{k(\delta_j,y_j(\theta,\delta)) - k(\delta_j^-,y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right] + R(\delta_j^r)\ .\end{aligned}$$

Since the participation constraints are binding, for all j, whenever $\delta_j = \delta^r$, we have $R(\delta_j^r) = \underline{\pi}_j$, for all j, and hence

$$\mathbb{E}_{(\theta,\delta)}\left[\left(k(\delta_j,y_j(\theta,\delta)) + \{k(\delta_j,y_j(\theta,\delta)) - k(\delta_j^-,y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right] + \underline{\pi}_j\ .$$

Consequently, a necessary condition for the implementability of $\{(q_i)_{i=1}^n, (y)_{j=1}^m\}$ is that

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\right)\right] \geq$$
$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j(\theta,\delta)) + \{k(\delta_j, y_j(\theta,\delta)) - k(\delta_j^-, y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\right)\right] .$$

Proof of Proposition 1.4.
Suppose that the condition in Proposition 1.4 holds. We need to show that we can construct a payment scheme satisfying all relevant constraints. Suppose first that the condition in Proposition 1.4 holds as an equality. Then we can choose a payment scheme that solves the relaxed problem, that we studied in the Proof of Proposition 1.3. To show this, we need to verify that the payment scheme which solves the relaxed problem in Proposition 1.3 satisfies not only the local downward incentive compatibility constraints, but all incentive compatibility constraints. This is a consequence of the Lemmata 1.2 and 1.9.
Since the given provision rule q_i satisfies the monotonicity constraints $q_i(\theta_{-i}, \theta_i^+, \delta) \geq q_i(\theta_{-i}, \theta_i, \delta)$ for all i and all l, Lemma 1.9 implies that all incentive compatibility constraints are binding. Lemma 1.3 implies that, for all i,

$$\mathbb{E}_{(\theta,\delta)}[t_i(\theta,\delta)] =$$
$$\mathbb{E}_{(\theta,\delta)}\left[\left(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] + T(\theta_i^0) .$$

Now choose

$$T(\theta_i^0) = \frac{1}{n}\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j(\theta,\delta)) + \{k(\delta_j, y_j(\theta,\delta)) - k(\delta_j^-, y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right) + \underline{\pi}_j\right]$$
$$- \frac{1}{n}\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right) - \underline{u}_i\right] ,$$

for all i. By assumption this is smaller or equal to zero, so that the interim participation constraints are satisfied, for all i. It remains to be shown that budget balance holds. This follows since, by construction,

$$
\begin{aligned}
& \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] \\
= \; & \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] + T(\theta_i^0) \\
= \; & \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] \\
& \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j,y_j(\theta,\delta)) + \{k(\delta_j,y_j(\theta,\delta)) - k(\delta_j^-,y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\right)\right] \\
& -\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left((v(\theta_i,q_i(\theta,\delta)) - \{v(\theta_i^+,q_i(\theta,\delta)) - v(\theta_i,q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\right)\right]
\end{aligned}
$$

Proof of Proposition 1.5.
Recall that at a solution to consumer transfer maximization

$$T_i(\theta^L) = \theta^L Q_i^n(\theta^L) \; ,$$

and

$$T_i(\theta^H) = \theta^H (Q_i^n(\theta^H) - Q_i^n(\theta^L)) + \theta^L Q_i^n(\theta^L) \; ,$$

which implies that

$$R^n = \frac{1}{n}\sum_{j=1}^{n}\{f(\theta^H)\theta^H(Q_i^n(\theta^H) - Q_i^n(\theta^L))\} + \theta^L Q_i^n(\theta^L) \; .$$

We can view the transfer maximization problem as consisting of $\theta^L Q_i^n(\theta^L)$ that every consumer has to pay and an incremental transfer if $f(\theta^H)\theta^H(Q_i^n(\theta^H) - Q_i^n(\theta^L))$ that applies only for consumers with a high valuation for the public good. Proposition 1.5 states that the revenue due to these incremental payments goes to zero as the number of consumers becomes large.
To proof this, we proceed in two steps:
Step 1. We first show that $lim_{n\to\infty} V^n = 0$, where

$$V^n := max_{Q^n()}\frac{1}{n}\sum_{i=1}^{n}\left(Q_i^n(\theta^H) - Q_i^n(\theta^L)\right) \; .$$

Fix θ'. The weight of $Q_n(\theta')$ in $\sum_{i=1}^{n}(Q_i^n(\theta^H) - Q_i^n(\theta^L))$ is given by

$$w^n(\theta') := m(\theta')f(\theta^H)^{m(\theta')-1}(1-f(\theta^H))^{n-m(\theta')} - (n-m(\theta'))f(\theta^H)^{m(\theta')}(1-f(\theta^H))^{n-1-m(\theta')} \; ,$$

where $m(\theta')$ is the number of individuals with $\theta'_i = \theta^H$.
Consequently, $Q^n(\theta')$ is chosen equal to 1 if $w^n(\theta) \geq 0$ and equal to 0 otherwise. Equivalently,

$$Q^n(\theta') = \begin{cases} 1, & \text{if } \frac{m(\theta')}{n} \geq f(\theta^H) , \\ 0 & \text{if } \frac{m(\theta')}{n} < f(\theta^H) . \end{cases}$$

Substituting these expression into V^n implies

$$V^n = \frac{1}{n} \sum_{x=f(\hat{\theta^H})n}^{n} \binom{n}{x} f(\theta^H)^x (1-f(\theta^H))^{n-x} - (n-x) f(\theta^H)^x (1-f(\theta^H))^{n-1-x} ,$$

where $f(\hat{\theta^H})n$ is the smallest integer larger than $nf(\theta^H)$. Equivalently,

$$\begin{aligned} V^n &= \frac{1}{f(\theta^H)(1-f(\theta^H))} \sum_{x=f(\hat{\theta^H})n}^{n} \binom{n}{x} f(\theta^H)^x (1-f(\theta^H))^{n-x} \left(\frac{x}{n} - f(\theta^H)\right) \\ &= \frac{1}{f(\theta^H)(1-f(\theta^H))} \sum_{x=f(\hat{\theta^H})n}^{n} prob\left(\frac{m(\theta)}{n} = x\right) \left(\frac{x}{n} - f(\theta^H)\right) . \end{aligned}$$

Note that $\frac{m(\theta)}{n} = \frac{1}{n}\sum_{i=1}^n z_i$, where $(z_i)_{i=1}^n$ is a collection function of *i.i.d.* random variables such that $z_i = 1$ if $\theta_i = \theta^H$ and $z_i = 0$ otherwise. By the strong Law of Large Numbers, for any $\epsilon > 0$,

$$lim_{n\to\infty} prob\left(\left|\frac{m(\theta)}{n} - f(\theta^H)\right| > \epsilon\right) = 0 .$$

Hence, $lim_{n\to\infty} V^n = 0$.
Step 2. Under any incentive compatible mechanism, $\frac{1}{n}\sum_{i=1}^n (Q_i^n(\theta^H) - Q_i^n(\theta^L))$ converges to zero. It follows from the incentive compatibility constraints that $Q_i^n(\theta^H) - Q_i^n(\theta^L) \geq 0$. This implies that $\frac{1}{n}\sum_{i=1}^n (Q_i^n(\theta^H) - Q_i^n(\theta^L)) \geq 0$. By *Step 1*, the upper bound on $\frac{1}{n}\sum_{i=1}^n (Q_i^n(\theta^H) - Q_i^n(\theta^L))$ converges to 0. Hence, $\frac{1}{n}\sum_{i=1}^n (Q_i^n(\theta^H) - Q_i^n(\theta^L))$ also converges to 0.

Proof of Proposition 1.6.
In order to proof that a further increase in the message set leads to Myerson and Satterthwaite (1983) impossibility result, we proceed as follows. In each round $k = 1, 2, \ldots$ we introduce more types for the buyer/the seller by placing new types at the center of every given sub-interval of $[\theta^L, \theta^H]$ and $[\delta^L, \delta^H]$. The partition into sub-intervals, we have in the beginning of round k is defined by the types set in all previous rounds. Thus, after round k we are facing a partition of $[\theta^L, \theta^H]$ into $2^{k-1} \cdot 2 = 2^k$ sub-intervals and the number of types is given by $\#\Theta = 2^k + 1, \#\Delta = 2^k + 1$. Since the first introducing-

round, we have added $2^k - 1$ types. In terms of the original types, the introduced types are given by

$1 = 1, \dots, 2^k$

$$\theta_k^l = \frac{1}{2^{k+1}} \left((\delta^H - \theta^L)(l - 2^k) + (l + 2^k) \right)$$

$$\delta_k^l = \frac{1}{2^{k+1}} \left((\delta^H - \theta^L) + 1 \right)$$

$$\theta_l^k - \delta^{k(2^k - l)} = \left(\delta^H - \theta^L \right) \frac{l - 2^k}{2^k} + \frac{l}{2^k}$$

Consider the Example and the mechanism that include $2^k + 1$ types. There is a social choice function which is efficient, implementable in Bayes-Nash equilibrium and yields non-negative material payoff for every type of individual if and only if

$$(\delta^H - \theta^L) \left(\sum_{l = \bar{l}_{kB}}^{1} \left[\frac{l - 2^k}{2^k} \frac{2^k - l + 1}{2^k + 1} \right] - (2^k - \bar{l}_{kB} + 1) \right) - \sum_{l = \bar{l}_{kB}}^{1} \frac{l}{2^k} \frac{2^k - l + 1}{2^k + 1} \geq 0 \ ,$$

where $l = \bar{l}_{kB}$ is the index of the buyer with valuation $\theta^{\bar{l}} := \min\{\theta : \theta > \delta^H\}$ is the lowest type, that does not lie in the interval $\delta^H - \theta^L$ in round k.
Remark: In every round $k = m$, the types introduced in round $\{1, \dots, m-1\}$ are given a new label, such that after round $k = m$ there does not exist any type θ_k^l, δ_k^l with $k \neq m$.
We assume that all types have equal probability $\frac{1}{2^k + 1}$.
Idea. Since the consumer perspective can be translated into the producer perspective by adjusting the indices, we only consider the consumer side to give an intuition for the stated result:

- To calculate the expected surplus in a certain round k (and thus to find out, if efficient implementation is possible), we sum up the gains from trade for all possible states of the world (combination of types), weighted with their probabilities.
- All types (expect the lowest and the highest) enter this sum positively and negatively. Positively, because buy the utility the considered type would get. And negatively, because each type is part of the information rent calculation of the type just below him. If these terms are equally high, the considered type does not enter the decision concerning efficient implementation.
- All θ_k^l for which $q(\theta_k^l, \cdot) = q(\theta_k^{(l-1)}, \cdot)$, for all δ_k^l (holding k fixed), cancel out. The reason is, that the probability mass of all positive and negative (hazard rate term of the type below θ_k^l) terms are equal. (The expected information rent effect equals the expected surplus effect.)

- If the condition above does not hold for the neighboring types, there remains a residuum by summing up the positive an negative terms, which means, that the considered type enters the expected surplus in a positive way.
- The weight by which a specific buyer θ_k^l enters the expected surplus is the probability of the state (θ_k^l, δ_k^l), for which $q(\theta_k^l, \delta_k^l) \neq q(\theta_k^{(l-1)}, \delta_k^l)$ multiplied with the sum over all probabilities of higher types. (This comes from the specific form of the hazard rate in discrete settings.)
- This is true for every buyer $\theta_k^l \leq \theta^{\bar{l}_{kB}}$.
- Adding up and rearranging all these terms provides the stated result.

Proof of Proposition 1.7.
Consider the Mailath and Postlewaite (1990) setup for $I = \{1, 2\}, \Theta = [\theta^L, \theta^H]$. We assume, that $\theta^L + \theta^H > 2c$. Then the efficient allocation rule is given by

$$q(\theta) = \begin{cases} 0, & \text{if} \quad \theta_1 + \theta_2 < 2c \ , \\ 1, & \text{if} \quad \theta_1 + \theta_2 > 2c \ . \end{cases}$$

Then g can be implemented, if and only if

$$\frac{1}{2} \sum_{i=1,2} \mathbb{E}_{(\theta)}[v(\theta_i) q(\theta)] \geq c \mathbb{E}_{(\theta)}[q(\theta)]$$

Or equivalently,

$$\int_{\theta_1} \int_{\theta_2} \frac{1}{2} \left(\theta_1 - \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} + \theta_2 - \frac{1 - F_2(\theta_2)}{f_2(\theta_2)} \right) q(\theta_1, \theta_2) f_1(\theta_1) f_2(\theta_2) d\theta_1 d\theta_2$$
$$\geq c \int_{\theta_1} \int_{\theta_2} q(\theta_1, \theta_2) f_1(\theta_1) f_2(\theta_2) d\theta_1 d\theta_2$$

$\Leftrightarrow$

$$\int_{\theta_1} \int_{max\{\theta_2^L, 2c-\theta_1\}}^{\theta_2^H} \frac{1}{2} \left(\theta_1 - \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} + \theta_2 - \frac{1 - F_2(\theta_2)}{f_2(\theta_2)} \right) f_1(\theta_1) f_2(\theta_2) d\theta_1 d\theta_2$$
$$\geq c \int_{\theta_1} \int_{max\{\theta_2^L, 2c-\theta_1\}}^{\theta_2^H} 1 f_1(\theta_1) f_2(\theta_2) d\theta_1 d\theta_2$$

The left-hand side of the inequality can be written as

$$\int_{\theta_1}\int_{max\{\theta_2^L,2c-\theta_1\}}^{\theta_2^H}\frac{1}{2}\left[\left(\theta_1-\frac{1-F_1(\theta_1)}{f_1(\theta_1)}+\theta_2\right)f(\theta_2)-1+F_2(\theta_2)\right]f_1(\theta_1)d\theta_1 d\theta_2$$

$$=\int_{\theta_1}\Bigg[\frac{1}{2}\left(\theta_1-\frac{1-F_1(\theta_1)}{f_1(\theta_1)}\right)F_2(\theta_2)\Big|_{max\{\theta_2^L,2c-\theta_1\}}^{\theta_2^H}+\theta_2^H$$

$$-max\{\theta_2^L,2c-\theta_1\}F_2(max\{\theta_2^L,2c-\theta_1\})-\int_{max\{\theta_2^L,2c-\theta_1\}}^{\theta_2^H}F_2(\theta_2)d\theta_2$$

$$-\theta_2^H+max\{\theta_2^L,2c-\theta_1\}+\int_{max\{\theta_2^L,2c-\theta_1\}}^{\theta_2^H}F_2(\theta_2)d\theta_2\Bigg]f_1(\theta_1)d\theta_1$$

$$=\int_{\theta_1}\frac{1}{2}\left((\theta_1-\frac{1-F_1(\theta_1)}{f_1(\theta_1)}+max\{\theta_2^L,2c-\theta_1\}\right)\left(1-F_2(max\{\theta_2^L,2c-\theta_1\})\right)f_1(\theta_1)d\theta_1$$

Case 1: $max\{\theta_2^L,2c-\theta_1\}=2c-\theta_1$.

$$\int_{\theta_2^L}^{2c-\theta_1^H}\frac{1}{2}\left(\theta_1-\frac{1-F_1(\theta_1)}{f_1(\theta_1)}+2c-\theta_1\right)\left(1-F_2(2c-\theta_1^H)\right)f_1(\theta_1)d\theta_1$$

$$=\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H}\left(-(1-F_1(\theta_1))(1-F_2(2c-\theta_1^H))+2cf_1(\theta_1)(1-F_2(2c-\theta_1^H))\right)d\theta_1$$

$$=\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H}F_2(2c-\theta_1^H)\left[1-2cf(\theta_1)-F_1(\theta_1)\right]d\theta_1$$

$$+\int_{\theta_2^L}^{2c-\theta_1^H}cf(\theta_1)d\theta_1-\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H}(1-F_1(\theta_1))d\theta_1$$

$$=\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H}\left[F_2(2c-\theta_1^H)-1\right](1-F_1(\theta_1))d\theta_1+cF_2(2c-\theta_1^H)$$

$$-\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H}F_2(2c-\theta_1^H)2cf(\theta_1)d\theta_1$$

Case 2: $max\{\theta_2^L,2c-\theta_1\}=\theta_2^L$.

$$\int_{2c-\theta_1^H}^{\theta_2^H}\frac{1}{2}\left(\theta_1-\frac{1-F_1(\theta_1)}{f_1(\theta_1)}+\theta_2^L\right)f_1(\theta_1)d\theta_1$$

$$=\frac{1}{2}\int_{2c-\theta_1^H}^{\theta_2^H}\left(\theta_1 f_1(\theta_1)-(1-F_1(\theta_1))+\theta_2^L f_1(\theta_1)\right)d\theta_1$$

$$=\frac{1}{2}\Bigg(\theta_1F_1(\theta_1)\Big|_{2c-\theta_1^H}^{\theta_2^H}-\int_{2c-\theta_1^H}^{\theta_2^H}F_1(\theta_1)d\theta_1$$

$$
\begin{aligned}
&- \int_{2c-\theta_1^H}^{\theta_2^H} (1 - F_1(\theta_1)) d\theta_1 + \theta_2^L F(\theta_1) \Big|_{2c-\theta_1^H}^{\theta_2^H} \Bigg) \\
&= \frac{1}{2} \left(\theta_2^H - (2c - \theta_1^H) F_1(2c - \theta_1^H) \right) - \int_{2c-\theta_1^H}^{\theta_2^H} 1 d\theta_1 + \theta_2^L (1 - F_1(2c - \theta_1^H)) \\
&= \frac{1}{2} \left(\theta_2^H - (2c - \theta_1^H) F_1(2c - \theta_1^H) - \theta_2^H + (2c - \theta_1^H) + \theta_2^L (1 - F_1(2c - \theta_1^H)) \right) \\
&= \frac{1}{2} (\theta_2^L - \theta_1^H)(1 - F_1(2c - \theta_1^H)) + c - cF_1(2c - \theta_1^H)
\end{aligned}
$$

Combining *Case 1* and *Case 2* the left-hand side of the inequality is given by

$$
\begin{aligned}
&\int_{\theta_1} \frac{1}{2} \left((\theta_1 - \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} + max\{\theta_2^L, 2c - \theta_1\} \right) \left(1 - F_2(max\{\theta_2^L, 2c - \theta_1\})\right) f_1(\theta_1) d\theta_1 \\
&= \frac{1}{2} \int_{\theta_2^L}^{2c-\theta_1^H} \left[F_2(2c - \theta_1^H) - 1 \right] (1 - F_1(\theta_1)) d\theta_1 + cF_2(2c - \theta_1^H) \\
&\quad - \frac{1}{2} \int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c - \theta_1^H) 2c f(\theta_1) d\theta_1 \\
&\quad + \frac{1}{2} (\theta_2^L - \theta_1^H)(1 - F_1(2c - \theta_1^H)) + c - cF_1(2c - \theta_1^H) \\
&= \frac{1}{2} \int_{\theta_2^L}^{2c-\theta_1^H} \left[F_2(2c - \theta_1^H) - 1 \right] (1 - F_1(\theta_1)) d\theta_1 + c \\
&\quad - \frac{1}{2} \int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c - \theta_1^H) 2c f(\theta_1) d\theta_1 + \frac{1}{2} (\theta_2^L - \theta_1^H)(1 - F_1(2c - \theta_1^H))
\end{aligned}
$$

The right-hand side of the inequality can be written as

$$
\begin{aligned}
&c \int_{\theta_1} \int_{max\{\theta_2^L, 2c-\theta_1\}}^{\theta_2^H} 1 f_1(\theta_1) f_2(\theta_2) d\theta_1 d\theta_2 \\
&= \int_{\theta_1} \left(1 - F_2(max\{\theta_2^L, 2c - \theta_1^H\})\right) f_1(\theta_1) d\theta_1 \\
&= c \int_{\theta_2^L}^{2c-\theta_1^H} \left(1 - F_2(2c - \theta_1^H)\right) f(\theta_1) d\theta_1 + \int_{2c-\theta_1^H}^{\theta_2^H} 1 f(\theta_1) d\theta_1 \\
&= c \left(\int_{\theta_2^L}^{\theta_2^H} 1 f_1(\theta_1) d\theta_1 - \int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c - \theta_1^H) f(\theta_1) d\theta_1 \right) \\
&= c \left(1 - \int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c - \theta_1^H) f(\theta_1) d\theta_1 \right)
\end{aligned}
$$

The inequality can hence be written as

$$
\begin{aligned}
&\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H} \left[F_2(2c-\theta_1^H)-1\right](1-F_1(\theta_1))d\theta_1 + c \\
&\quad -\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c-\theta_1^H)2cf(\theta_1)d\theta_1 + \frac{1}{2}(\theta_2^L-\theta_1^H)(1-F_1(2c-\theta_1^H)) \\
&\geq c\left(1-\int_{\theta_2^L}^{2c-\theta_1^H} F_2(2c-\theta_1^H)f(\theta_1)d\theta_1\right)
\end{aligned}
$$

$\Leftrightarrow$

$$
\begin{aligned}
&\frac{1}{2}\int_{\theta_2^L}^{2c-\theta_1^H} \underbrace{\left[F_2(2c-\theta_1^H)-1\right]}_{<0}(1-F_1(\theta_1))d\theta_1 + \frac{1}{2}\underbrace{(\theta_2^L-\theta_1^H)}_{<0}(1-F_1(2c-\theta_1^H)) \\
&\geq 0
\end{aligned}
$$

The inequality cannot be fulfilled. The public good can therefore not be provided efficiently.

Proof of Proposition 1.8.
Reminder: f_D, F_D are the density-, distribution function in the discrete environment. For every individual, they are defined as follows:

$$f_D(\theta) = \int_{\frac{\theta^l+\theta^{l-1}}{2}}^{\frac{\theta^{l+1}+\theta^l}{2}} f(\theta^l)d\theta^l \ , \qquad \theta^l \in (\theta^0, \theta^s)$$

$$
\begin{aligned}
f_D(\theta^0) &= \int_0^{\frac{\theta^1+\theta^0}{2}} f(\theta^l)d\theta^l \\
f_D(\theta^s) &= \int_{\frac{\theta^s+\theta^{s-1}}{2}}^{\theta^s} f(\theta^l)d\theta^l
\end{aligned}
$$

Without loss of generality (for $I < \infty$), we consider a situation with one individual on each side we take δ as given and calculate the ex ante expected payment. All arguments also apply in case of more than one agent and taking expectation over (θ, δ). The ex ante expected payment is given by the following equation

$$\mathbb{E}_\theta\left[\left(v(\theta^l, q^*(\theta,\delta))\right] - \sum_{l=1}^{s} f(\theta^l)\{v(\theta^{l+1}, q^*(\theta,\delta)) - v(\theta^l, q^*(\theta,\delta))\}\frac{1-F_D(\theta^l)}{f_D(\theta^l)}\right) \ .$$

We neglect the first term for the moment and consider the second:

$$\begin{aligned}
&\sum_l f(\theta^l)\{v(\theta^{l+1}, q^*(\theta,\delta)) - v(\theta^l, q^*(\theta,\delta))\}\frac{1-F_D(\theta^l)}{f_D(\theta^l)} \\
&= \sum_l \{v(\theta^{l+1}, q^*(\theta,\delta)) - v(\theta^l, q^*(\theta,\delta))\}(1-F_D(\theta^l)) \\
&= \sum_l \frac{v(\theta^{l+1}, q^*(\theta,\delta)) - v(\theta^l, q^*(\theta,\delta))}{\theta^{l+1}-\theta^l}(\theta^{l+1}-\theta^l)(1-F_D(\theta^l)) \ .
\end{aligned}$$

Adding types in every round k, yields that $|\theta^{l+1}-\theta^l| \to 0$. Thus the difference quotient of the value function converges to the partial derivative of v with respect to θ_l.

$$\begin{aligned}
&\lim_{|\theta^{l+1}-\theta^l|\to 0} \sum_l \frac{v(\theta^{l+1}, q^*(\theta,\delta)) - v(\theta^l, q^*(\theta,\delta))}{\theta^{l+1}-\theta^l}(\theta^{l+1}-\theta^l)(1-F_D(\theta^l)) \\
&= \sum_l \frac{\partial v}{\partial \theta^l}\Big|_{\theta^l}(1-F_D(\theta^l)) \ .
\end{aligned}$$

Since v is monotone we can change summation and take the infinum/supremum. The limit of the upper- and the lower-sum of the above function coincide and $F_D(\theta^l) \to F(\theta^l)$, thus the integral exists and is given by[20]

$$\begin{aligned}
&\lim_{|\theta^{l+1}-\theta^l|\to 0} \sum_l \inf_z \left(\frac{\partial v}{\partial \theta^l}\Big|_z(1-F_D(z))\right) = \lim_{|\theta^{l+1}-\theta^l|\to 0} \sum_l \sup_z \left(\frac{\partial v}{\partial \theta^l}\Big|_z(1-F_D(z))\right) \\
&= \int_{\theta_0}^{\theta^s} \frac{\partial v}{\partial \theta^l}\Big|_z (1-F(z))dz \ .
\end{aligned}$$

Using integration by parts and Fubini's Theorem (to change the order of integration), we get

$$\begin{aligned}
\int_{\theta^0}^{\theta^s} \frac{\partial v}{\partial \theta^l}\Big|_s (1-F(s))dr &= \int_{\theta^0}^{\theta^s} \frac{\partial v}{\partial \theta^i}\Big|_x (1-F(x))dx \\
&= \int_{\theta^0}^{\theta^s} \frac{\partial v}{\partial \theta^l}\Big|_x \int_x^{\theta^s} f(z)dzdx \\
&= \int_{\theta^0}^{\theta^s} \int_x^{\theta^s} \frac{\partial v}{\partial \theta^l}\Big|_x f(z)dzdx \\
&= \int_{\theta^0}^{\theta^s} \int_{\theta^s}^{z} \frac{\partial v}{\partial \theta^l}\Big|_x f(z)dxdz \\
&= \int_{\theta^0}^{\theta^s} \int_{\theta^s}^{z} \frac{\partial v}{\partial \theta^l}\Big|_x dx f(z)dz \ .
\end{aligned}$$

[20]Reminder: We summarize over all possible types $l = 0, \ldots, s$. This is not a summation over agents.

For the neglected term $\mathbb{E}_\theta\left[v(\theta^l, q^*(\theta,\delta))\right] = \sum_l f_D(\theta^l)v(\theta^l, q^*(\theta,\delta))$, we consider the limes inferior and the limes superior and find, that it converges:

$$\begin{aligned}&\lim_{|\theta^{l+1}-\theta^l|\to 0}\sum_l \inf_{\theta^l}\left(v(\theta^l, q^*(\theta^l,\delta))\right) = \lim_{|\theta^{l+1}-\theta^l|\to 0}\sum_l \sup_{\theta^l}\left(v(\theta^l, q^*(\theta^l,\delta))\right)\\ &= \int_{\theta^0}^{\theta^s} v(z, q^*(z,\delta))f(z)dz\ .\end{aligned}$$

Combining the converged terms, this yields $\int_{\theta^0}^{\theta^s}\left(v(z,q^*(z,\delta)) - \int_{\theta^s}^{z}\frac{\partial v}{\partial\theta^l}\Big|_x dx\right)f(z)dz = \mathbb{E}_\theta[t(\theta,\delta)]$, which gives the claimed result.

Proof of Corollary 1.1.

Let us assume for a moment that the mechanism that maximizes expected surplus subject to (IC_F) and (PC_F) satisfies the monotonicity constraint $y_j(\theta,\delta_{-j},\delta_j^l) \geq y_j(\theta,\delta_{-j},\delta_j^{l+1})$, for all l. This will be verified below.

A necessary condition for the maximization of S is that consumers' transfers need to be higher than firms' revenues. The budget constraint (1.6) needs to hold with equality, otherwise it was possible to decrease transfers without violating any constraints of the surplus maximization problem. Hence,

$$S(\theta,\delta) := \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^n (v(\theta_i, q_i(\theta,\delta)) - \sum_{j=1}^m k(\delta_j, y_j(\theta,\delta))\right]\ .$$

Now suppose that the solution to this problem involves overproduction,

$$\sum_{i=1}^n q_i(\theta,\delta) \leq \sum_{j=1}^m y_j(\theta,\delta)\ .$$

Then increasing $\sum_{i=1}^n q_i(\theta,\delta)$ involves no costs, i.e. firm profits remain unaffected, but increases consumer surplus as $\mathbb{E}_{(\theta,\delta)}[\sum_{i=1}^n v(\theta_i, q_i(\theta,\delta))]$ goes up. This is a contradiction to the assumption that the optimum involves underproduction. Hence, we need that $\sum_{i=1}^n q_i(\theta,\delta) = \sum_{j=1}^m y_j(\theta,\delta)$.

We can therefore once more rewrite the problem of choosing an optimal provision rule: Choose $(q_i)_{i=1}^n, (y_j)_{j=1}^m$ in order to maximize

$$S(\theta,\delta) := \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^n v(\theta_i, q_i(\theta,\delta)) - \sum_{j=1}^m k(\delta_j, y_j(\theta,\delta))\right]\ ,$$

subject to $\sum_{i=1}^n q_i(\theta,\delta) = \sum_{j=1}^m y_j(\theta,\delta)$. The solution to that problem is such that the following first order condition is satisfied:

$$\sum_{i=1}^{n} v_2(\theta_i, q_i^*(\theta,\delta)) = \sum_{j=1}^{m} k_2(\delta_j, y_j^*(\theta,\delta)) \ .$$

The first order condition implies that for every δ, $q_i(\theta_{-i}, \theta_i^l) \geq q_i(\theta_{-i}, \theta_i^{l-1})$ and for every $y_j(\delta_{-j}, \delta_j^l) \geq y_j(\delta_{-j}, \delta_j^{l+1})$. This implies that the monotonicity conditions, for all l, are satisfied.

Now we construct the expected transfers of consumers such that all local downward incentive compatibility constraints are binding and choose $T(\theta_i^0)_{i=1}^n$ such that

$$\sum_{i=1}^{n} T(\theta_i^0) = \mathbb{E}_{(\theta,\delta)}\Big[\sum_{i=1}^{n}\Big(v(\theta_i, q_i^*(\theta,\delta)) - \{v(\theta_i^+, q_i^*(\theta,\delta)) - v(\theta_i, q_i^*(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\Big) - \sum_{j=1}^{m}\Big(k(\delta_j, y_j^*(\theta,\delta)) + \{k(\delta_j, y_j^*(\theta,\delta)) - k(\delta_j^-, y_j^*(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\Big)\Big] \ .$$

It follows from Lemma 1.2 that for all consumers incentive compatibility constraints are satisfied. Also it follows from Proposition 1.4, that the expected transfers are given by

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] = \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j^*(\theta,\delta)) + \{k(\delta_j, y_j^*(\theta,\delta)) - k(\delta_j^-, y_j^*(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\right)\right] - \sum_{i=1}^{n}\underline{u}_i \ .$$

It follows as well from Proposition 1.4 that maximal revenue that can be extracted from consumers are equal to

$$\mathbb{E}_{(\theta,\delta)}\Big[\sum_{i=1}^{n}\Big(v(\theta_i, q^*(\theta,\delta)) - \{v(\theta_i^+, q_i^*(\theta,\delta)) - v(\theta_i, q_i^*(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\Big)\Big] \ .$$

Consequently, a necessary condition for the implementability of $(q_i^*)_{i=1}^n, (y_j^*)_{j=1}^m$ is that

$$\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\left(v(\theta_i, q_i^*(\theta,\delta)) - \{v(\theta_i^+, q_i^*(\theta,\delta)) - v(\theta_i, q_i^*(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\right)\right] - \sum_{i=1}^{n}\underline{u}_i \geq \mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\left(k(\delta_j, y_j^*(\theta,\delta)) + \{k(\delta_j, y_j^*(\theta,\delta)) - k(\delta_j^-, y_j^*(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\right)\right] + \sum_{j=1}^{m}\underline{\pi}_j \ .$$

Sufficiency of this condition can be shown by using once more, the construction in the

Proof of Proposition 1.4. If the condition above holds and we let for all i,

$$T(\theta_i^0) - \Big(E\Big[\sum_{i=1}^{n}\Big(v(\theta_i, q_i^*(\theta,\delta)) - \{v(\theta_i^+, q_i^*(\theta,\delta)) - v(\theta_i, q_i^*(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\Big)$$
$$-\sum_{j=1}^{m}\Big(k(\delta_j, y_j^*(\theta,\delta)) + \{k(\delta_j, y_j^*(\theta,\delta)) - k(\delta_j^-, y_j^*(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\Big)\Big]\Big) \geq 0\ ,$$

we obtain a mechanism that achieves $(q_i^*)_{i=1}^n, (y_j^*)_{j=1}^m$, satisfies all relevant constraints.

Proof of Corollary 1.2.
Suppose that the constraint in Corollary 1.1 is violated, then the inequality constraint in Proposition 1.2 is binding, and the optimal provision rule maximizes the following Lagrangian

$$\mathcal{L} = \mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}(v(\theta_i, q(\theta,\delta)) - k(\delta_j, y_j(\theta,\delta)))\right]$$
$$+\lambda\Bigg(\mathbb{E}_{(\theta,\delta)}\left[\sum_{i=1}^{n}\Big(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)} - \underline{u}_i\Big)\right]$$
$$-\mathbb{E}_{(\theta,\delta)}\left[\sum_{j=1}^{m}\Big(k(\delta_j, y_j(\theta,\delta)) + \{k(\delta_j, y_j(\theta,\delta)) - k(\delta_j^-, y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)} + \underline{\pi}_j\Big)\right]\Bigg)\ ,$$

where λ is the Lagrangian multiplier which, at a solution to this maximization problem, has to be strictly positive, $\lambda > 0$.
To complete the proof it remains to be shown that a solution to the maximization problem satisfies the monotonicity constraints.
To see that $q_i(\theta_{-i}^+, \theta_i) \geq q_i(\theta_{-i}, \theta_i)$, for all i and all l holds, note that the monotone hazard rate assumption implies that

$$\mathbb{E}_{(\theta,\delta)}\left[\Big(v(\theta_i, q_i(\theta,\delta)) - \{v(\theta_i^+, q_i(\theta,\delta)) - v(\theta_i, q_i(\theta,\delta))\}\frac{1-F(\theta_i)}{f(\theta_i)}\Big)\right] \geq 0\ ,$$

is an increasing function. Consequently, the solution to the maximization problem is such that $q_i(\theta_{-i}^+, \theta_i, \delta) \geq q_i(\theta_{-i}, \theta_i, \delta)$, for all i, θ_{-i} and δ.
To see that $y_j(\theta, \delta_{-j}, \delta_j^-) \geq y_j(\theta, \delta_{-j}, \delta_j)$, for all j , note that

$$\mathbb{E}_{(\theta,\delta)}\left[\Big(k(\delta_j, y_j^*(\theta,\delta)) + \{k(\delta_j, y_j(\theta,\delta)) - k(\delta_-, y_j(\theta,\delta))\}\frac{P(\delta_j)}{p(\delta_j)}\Big)\right] \geq 0\ ,$$

also is an increasing function. This implies that, for all j, δ_{-j} and θ, $y_j(\theta, \delta_{-j}, \delta_j^-) \geq y_j(\theta, \delta_{-j}, \delta_j)$.

Appendix 1.C Applications

Here, we present a further application that highlight the difference between discrete and continuous type settings. For the partnership dissolution framework of Cramton, Gibbons, and Klemperer (1987), we demonstrate that irrespectively of the original distribution of shares in the partnership, this partnership can be dissolved efficiently for examples with discrete types.

Partnership dissolution, Cramton, Gibbons, and Klemperer (1987). There are $I = \{1, 2\}$ consumers, also referred to as partners, and no producers. They form a partnership in which any one agent i initially holds a share $e_i \in [0, 1]$, with $e_1 + e_2 = 1$. The allocation problem is to change the ownership structure. Let s_i be agent i's share in the partnership after the reassignment of shares. Let t_i be the monetary payment of i, which is positive if i has to compensate others for receiving their shares and is negative if i sells some of her shares to other partners. Partner i evaluates this outcome according to the utility function $\theta_i \; s_i - t_i$. We can relate this setup to the general framework developed in Section 1.4 by defining $q_i = s_i - e_i$ as the change of the shares held by agent i. Partner i's utility gain from the change of the ownership structure can then be written as $\theta_i \; q_i - t_i$. A social choice function consists of a collection of consumption functions $q_i : \Theta^n \to \mathbb{R}$, $i \in I$, so that, for all θ,

$$q_1(\theta) + q_1(\theta) = 0 \quad \text{and,} \quad \text{for all } i, \; -e_i \leq q_i(\theta) \leq 1 - e_i \; . \tag{1.15}$$

The partners have to agree unanimously on the new ownership structure so that $\underline{u}_i = 0$, for all i. The surplus that is generated by the change of the ownership structure is, again, given by $\mathbb{E}_\theta \left[\theta_1 q_1(\theta) + \theta_2 q_2(\theta)\right]$.

A key insight by Cramton, Gibbons, and Klemperer (1987) is that the specification of the initial ownership structure, i.e. the choice of $e = (e_1, e_2)$, has an influence on the possibility to dissolve a partnership in a surplus-maximizing way. Below, we will show that our discrete type specification allows us to communicate this insight in a very simply way, without having to invoke all the calculus that an analysis with an atomless distribution would require.

In the following we assume that there are two types per individual, i.e. $\Theta_1 = \{\theta_1^L, \theta_1^H\}$ and $\Theta_2 = \{\theta_2^L, \theta_2^H\}$. We denote the probability of the event $\theta_i = \theta_i^L$ by f_i^L and $f_i^H := 1 - f_i^L$. For ease of notation we define the interim expected change of share in partnership as:

$$Q_1(\theta_1^L) := f_1^L q_1(\theta_1^L, \theta_2^L) + f_1^H q_1(\theta_1^L, \theta_2^H) \; ,$$

and analogously $Q_1(\theta_1^H)$, $Q_2(\theta_2^L)$ and $Q_2(\theta_2^H)$

Proposition 1.9. *Suppose that q_1 and q_2 are such that*

$$Q_1(\theta_1^L) \leq Q_1(\theta_1^H) \leq 0, \quad \textit{and} \quad Q_2(\theta_2^L) \leq Q_2(\theta_2^H) \leq 0 \ ,$$

then there exists (t_1, t_2) so that (q_1, q_2, t_1, t_2) satisfies incentive compatibility, voluntary participation and expected budget balance if and only if

$$\mathbb{E}_\theta[\theta_1 q_1(\theta_1, \theta_2) + \theta_2 q_2(\theta_1, \theta_2)] - f_1^L(\theta_1^H - \theta_1^L)Q_1(\theta_1^H) - f_2^H(\theta_2^H - \theta_2^L)Q_2(\theta_2^L) \geq 0 \ .$$

In a discrete type version of Cramton, Gibbons, and Klemperer (1987) an efficient dissolution of partnership is possible, even though the initial ownership is such that one party owns everything and the other party nothing. This is different under a continuous distribution of types. The possibility result relates to Observation 1.1. If initial shares are distributed more equally between two parties, efficient dissolution of partnership is possible with discrete and continuous distribution of types, as the following Proposition highlights.

Proposition 1.10. *Suppose that q_1 and q_2 are such that*

$$Q_1(\theta_1^L) \leq 0 \leq Q_1(\theta_1^H), \quad \textit{and} \quad Q_2(\theta_2^L) \geq 0 \geq Q_2(\theta_2^H) \ ,$$

then there exists (t_1, t_2) so that (q_1, q_2, t_1, t_2) satisfies incentive compatibility, voluntary participation and expected budget balance if and only if

$$\mathbb{E}_\theta[\theta_1 q_1(\theta_1, \theta_2) + \theta_2 q_2(\theta_1, \theta_2)] \geq 0 \ .$$

Proofs of Application.

Proof of Proposition 1.9.
For the given allocation function q the incentive compatibility and participation constraints for individual i are given by

$$\theta_i^L Q_i(\theta_i^L) - T_i(\theta_i^L) \geq \theta_i^L Q_i(\theta_i^H) - T_i(\theta_i^H) \tag{1.16}$$

$$\theta_i^H Q_i(\theta_i^H) - T_i(\theta_i^H) \geq \theta_i^H Q_i(\theta_i^L) - T_i(\theta_i^L) \tag{1.17}$$

$$\theta_i^L Q_i(\theta_i^L) - T_i(\theta_i^L) \geq 0 \tag{1.18}$$

$$\theta_i^H Q_i(\theta_i^H) - T_i(\theta_i^H) \geq 0 \ . \tag{1.19}$$

Additional we have the expected budget balance condition:

$$\mathbb{E}_{(\theta)}\left[t_1(\theta_1, \theta_2) + t_2(\theta_1, \theta_2)\right] = 0$$

For standard arguments, we assume, that the participation constraint for the worst-off type and the incentive compatibility constraint for the best-off are binding and solve the relaxed problem under these conditions. Subsequently we show, that the found solution fulfills the neglected conditions. Since $Q_1(\theta_{1j}) \leq 0, j \in L, H$ we refer to individual 1 as the seller and individual 2 as the buyer. Thus equations (1.17) and (1.18) are binding for individual 2 and equations (1.16) and (1.19) are binding for individual 1.
Rearranging these conditions yields

$$T_2(\theta_2^L) = \theta_2^L Q_2(\theta_2^L)$$
$$T_1(\theta_1^H) = \theta_1^H Q_1(\theta_1^H)$$
$$T_2(\theta_2^H) = \theta_2^H Q_2(\theta_2^H) - (\theta_2^H - \theta_2^L) Q_2(\theta_2^L)$$
$$T_1(\theta_1^L) = \theta_1^L Q_1(\theta_1^L) - (\theta_1^L - \theta_1^H) Q_1(\theta_1^H)$$

Plugging these transfers into the neglected conditions:

$$0 \geq \theta_2^L Q_2(\theta_2^H) - \left(\theta_2^H Q_2(\theta_2^H) - (\theta_2^H - \theta_2^L) Q_2(\theta_2^L)\right) \Leftrightarrow Q_2(\theta_2^H) \geq Q_2(\theta_2^L) \tag{1.20}$$

$$\theta_2^H Q_2(\theta_2^H) - \theta_2^H Q_2(\theta_2^H) - (\theta_2^H - \theta_2^L) Q_2(\theta_2^L) \geq 0 \; . \tag{1.21}$$

Equation (1.21) always holds. The constraints for player 1 are analogous. We check now the budget balance condition

$$\begin{aligned} & f_1^H \left(\theta_1^H Q_1(\theta_1^H)\right) + f_1^L \left(\theta_1^L Q_1(\theta_1^L) - (\theta_1^L - \theta_1^H) Q_1(\theta_1^H)\right) \\ = & - \left(f_2^L(\theta_2^L Q_2(\theta_2^L)) + f_2^H(\theta_2^H Q_2(\theta_2^H) - (\theta_2^H - \theta_2^L) Q_2(\theta_2^L))\right) \; . \end{aligned}$$

Thus, for $Q_i(\theta_i^H) \geq Q_i(\theta_i^L)$ we found the desired transfers, such that $t_1(\theta_1, \theta_2)$, $t_2(\theta_1, \theta_2)$, $q_1(\theta_1, \theta_2)$, $q_2(\theta_1, \theta_2)$ fulfill incentive compatibility constraints, participation constraints, if and only if

$$\mathbb{E}_\theta[\theta_1 q_1(\theta_1, \theta_2) + \theta_2 q_2(\theta_1, \theta_2)] - f_1^L(\theta_1^H - \theta_1^L) Q_1(\theta_1^H) - f_2^H(\theta_2^H - \theta_2^L) Q_2(\theta_2^L) \geq 0 \; .$$

Proof of Proposition 1.10.
We solve the relaxed problem and check, whether the solution fulfill the neglected conditions. Incentive compatibility and participation constraints are given by equations (1.16) - (1.19).
We assume that the participation constraints are binding. Thus

$$T_i(\theta_i^L) = \theta_i^L Q_i(\theta_i^L)$$
$$T_i(\theta_1^H) = \theta_i^H Q_i(\theta_i^H)$$

Check for the incentive compatibility constraints:

$$0 \geq \theta_i^L Q_i(\theta_i^H) - \theta_i^H Q_i(\theta_i^H)$$
$$0 \geq \theta_i^H Q_i(\theta_i^L) - \theta_i^L Q_i(\theta_i^L) \ .$$

Since $Q_i(\theta_i^H) \geq 0$ and $Q_i(\theta_i^L) \leq 0$, both incentive compatibility constraints hold. To check budget balance, we plug the transfers into the budget balance condition and get the desired result.

Appendix 1.D From discrete to continuous for the firm side

Production Economy. We define a production *economy* (J, Δ, p, y) as the vector consisting of producers $j \in J$, the Cartesian product of type sets (one type set for every firm j) $\Delta = \Delta^1 \times \cdots \times \Delta^m$ with the corresponding density functions $p^j : \Delta_j \to [0, 1]$, where $p := (p_0, \ldots, p_m)$, and the allocation rule $y : \Delta \to \mathbb{R}_+$, that maps reported types into allocations. Thus, a sequence of production economies is given by $(J_k, \Delta_k, p_k, y_k)_{k\in\mathbb{N}}$, where J_k is a sequence of producer sets, Δ_k a sequence of type set-products, p_k a sequence of density-function vectors and y_k a sequence of allocation functions. We introduce some additional definitions and notations:
Producer. We keep the set $J = 1, \ldots, m_k$ be the set of consumers in round k. For ease of notation, we write J instead of J_k and $J_{\infty k}$ in every element of the sequence of production economies.
Type sets. An element Δ_l of the sequence $(\Delta_k)_{k\in\mathbb{N}}$ is given by the Cartesian product over the type sets of the firms in round k. Since we assume symmetric firms, every firm $\kappa \in J$ has the same type set $\Delta_{k\kappa}$. Thus $\Delta_k = (\Delta_{k\kappa})^{m_k}$
Density Functions. The sequence $(p_k)_{k\in\mathbb{N}}$ consists of $p_k = (p_{k1}, \ldots, p_{km_k})$, where p_{kj} is the density function over a finite type set, namely Δ_{kj}, for all $j < m_k < \infty, j \in J$.
*Allocation Functions.*The sequence of allocations functions is given by $(y_k)_{k\in\mathbb{N}} : (\Delta_k)_{k\in\mathbb{N}} \to (x_k)_{k\in\mathbb{N}}, x_k \in \mathbb{R}$, where $y_k : \Delta_k \to \mathbb{R}$. *Remark*: The definitions of convergence are analogously to the consumption side. The revenues that we receive for a mechanism (y, r) in a continuous environment, that fulfills incentive compatibility constraints and participation constraints can be found in equation (1.10)

Proposition 1.11. *Let $(J, \Delta_k, p_k, y_k)_{k\in\mathbb{N}} \to (J, \Delta_\infty, p_\infty, y_\infty)$, for $k \to \infty$. Then it holds, that for every producer $j \in J$: For every $\epsilon > 0 \; \exists \quad K : \forall \; k \geq K$*

$$\Bigg| \mathbb{E}_{(\delta_{kj})} \left[\left(k(\delta_{kj}, y_{kj}(\theta, \delta)) + \{k(\delta_{kj}, y_{kj}(\theta, \delta)) - k(\delta^-_{k(kj)}, y_{kj}(\theta, \delta))\} \frac{P_k(\delta_{kj})}{p_k(\delta_{kj})} \right) \right] - \mathbb{E}_{(\delta_{\infty j})} \left[k(\delta_{\infty j}, y_{\infty j}(\theta, \delta)) + k_1(\delta_{\infty j}, y_{\infty j}(\theta, \delta)) \frac{P_\infty(\delta_{\infty j})}{p_\infty(\delta_{\infty j})} d\delta_{\infty j} \right] \Bigg| < \epsilon \, .$$

Proof of Proposition 1.11.
The arguments are similar to the ones used in Proposition 1.8 and therefore left to the reader.

Appendix 1.E Proof of Observations

Proof of Observation 1.1.
In the direct mechanism, there are 4 states of the economy, namely: (θ^H, δ^H), (θ^H, δ^L), (θ^L, δ^H) and (θ^L, δ^L). Since $q(\theta^L, \delta^H) = 0$, straightforward computations yield

$$\begin{aligned}
&(\theta_1^H, \delta_1^H) : f^H f^L \left[\theta^H - \delta^H + (\delta^H - \delta^L)\frac{f^H}{f^L}\right] \\
&(\theta_1^H, \delta_1^L) : + f^H f^H \left[\theta^H - \delta^L\right] \\
&(\theta_1^L, \delta_1^L) : + f^L f^H \left[\theta^L - (\theta^H - \theta^L)\frac{f^H}{f^L} - \delta^L\right] \geq 0
\end{aligned}$$

Hence, implementation of the social choice function is possible if and only if

$$f^L \geq \frac{\delta^H - \theta^L}{\theta^H - \delta^L} \ .$$

Proof of Observation 1.2.
The expected surplus is

$$\begin{aligned}
S() =& f^H f^L \left[\theta^H - \delta^H\right] + f^H f^H \left[\theta^H - \delta^L\right] + f^L f^H \left[\theta^L - \delta^L\right] \\
=& f^H f^L \left[\theta^H - \delta^H + \theta^L - \delta^L\right] + f^H f^H \left[\theta^H - \delta^L\right] \\
=& f^H f^L \left[1 - (\delta^H - \theta^L)\right] + f^H f^H 1 \\
=& f^H f^L \left[1 - d\right] + f^H f^H
\end{aligned}$$

Hence $\frac{\partial S()}{\partial d} < 0$.
The expected information rents are

$$\begin{aligned}
IR() =& f^H f^L \left[\delta^H - \delta^L + \theta^H - \theta^L\right] \\
=& f^H f^L \left[1 + \delta^H - \theta^L\right] \\
=& f^H f^L \left[1 + d\right]
\end{aligned}$$

Hence $\frac{\partial IR()}{\partial d} > 0$.

Proof of Observation 1.3.
From the proof of Observation 1.2, we know that

$$S() = (1 - f^L) f^L \left[1 - d\right] + (1 - f^L)^2 \ .$$

Therefore, whenever $\theta^H - \delta^H < \frac{1}{2}$,
$\frac{\partial S()}{\partial f^L} < 0$.

The expected information rents are

$$IR() = f^H f^L [1-d] \ .$$

So that $\frac{\partial IR()}{\partial f^L} < 0$.

Proof of Observation 1.4
Consider the Mailath and Postlewaite (1990) setup for $I = \{1,2\}, \Theta_i = \Theta = \{\theta^L, \theta^H\}$. We assume, that $0.5\left(\theta^L + \theta^H\right) > c$. (For ease of notation: $f(\theta^L) = f^L$) Then g can be implemented efficiently, if and only if

$$\frac{1}{2}\sum_{i=1,2} \mathbb{E}_{(\theta)}[v(\theta_i)g(\theta)] \geq c\mathbb{E}_{(\theta)}[g(\theta)] \ .$$

Since trade takes place, if at least one consumer has a high valuation, and g is either 0 or 1, the right-hand side of the equation equals

$$\begin{aligned} c\mathbb{E}_{(\theta)}[g(\theta)] =& c(1 - Prob(\text{both have a low valuation})) \\ &= c\left(1-(f^L)^2\right) = c\left((1+f^L)(1-f^L)\right) = c\left((1+f^L)f^H\right) \ . \end{aligned}$$

The left-hand side is given by

$$\begin{aligned} &\frac{1}{2}\sum_{i=1,2}\mathbb{E}_{(\theta)}[v(\theta_i)g(\theta)] \\ &\quad = \frac{1}{2}(f^H)^2\left[2\theta^H\right] + \frac{1}{2}2f^H\ f^L\left[\theta^L - \frac{1-F(\theta^L)}{f^L}(\theta^H-\theta^L) + \theta^H\right] \\ &\quad = f^H\left(\theta^L + f^L\theta^H\right) \ . \end{aligned}$$

This yields that efficient implementation is possible, if and only if

$$\theta^L + f^L\theta^H \geq (1-f^L)c \Leftrightarrow f^L \geq \frac{c-\theta^L}{\theta^H - c}$$

Proof of Observation 1.5.

$$MS = -\left(\underbrace{\left[\theta^H - c - f_L(c-\theta^L)\right]}_{S(\cdot)} - \underbrace{(1-f^L)(\theta^H-\theta^L)}_{IR(\cdot)}\right) \leq 0 \ .$$

i) If f^L is fixed, then $\frac{\partial S()}{\partial c} < 0$.
ii) If c is fixed, then $\frac{\partial S()}{\partial c} < 0$, and $\frac{\partial IR()}{\partial c} < 0$.

Proof of Observation 1.6.
Consider the mechanism, in which the third type θ^M and δ^M for each player is added by cutting the existing intervals $[\theta^H, \theta^L]$ and $[\delta^H, \delta^L]$ into halves and set

$$\delta^M = \frac{1}{2}\left(\delta^H - \theta^L + \frac{1-(\delta^H - \theta^L)}{2}\right) \ ,$$

and

$$\theta^M = \frac{1}{2}\left(\delta^H - \theta^L + \frac{1-(\delta^H - \theta^L)}{2}\right) + \frac{1-(\delta^H - \theta^L)}{2} \ .$$

The mechanism works like the mechanism with outcome function f. We assume, that the new middle type has probability, f_1^M.
Efficient implementation is possible, whenever Corollary 1.1 applies.
If an additional type is introduced for the buyer and the seller, there are 9 states of the economy. Trade is inefficient when the low valuation buyer faces the high cost seller.

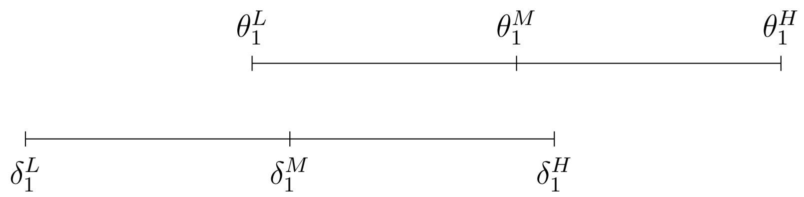
$$\begin{aligned}
(\theta^H, \delta^H) &: f_1^H f_1^L \left[\theta^H - \left(\delta^H + (\delta^H - \delta^M)\frac{f_1^H + f_1^M}{f_1^L}\right)\right] \\
(\theta^H, \delta^M) &: f_1^H f_1^M \left[\theta^H - \left(\delta^M + (\delta^M - \delta^L)\frac{f_1^H}{f_1^M}\right)\right] \\
(\theta^H, \delta^L) &: f_1^H f_1^H \left[\theta^H - \delta^L\right] \\
(\theta^M, \delta^H) &: f_1^M f_1^H \left[\theta^M - (\theta^H - \theta^M)\frac{f_1^H}{f_1^M} - \left(\delta^H + (\delta^H - \delta^M)\frac{f_1^H + f_1^M}{f_1^L}\right)\right] \\
(\theta^M, \delta^M) &: f_1^M f_1^M \left[\theta^M - (\theta^H - \theta^M)\frac{f_1^H}{f_1^M} - \left(\delta^M + (\delta^H - \delta^M)\frac{f_1^H}{f_1^M}\right)\right] \\
(\theta^M, \delta^L) &: f_1^M f_1^L \left[\theta^M - (\theta^H - \theta^M)\frac{f_1^H}{f_1^M} - \delta^L\right] \\
(\theta^L, \delta^M) &: f_1^L f_1^M \left[\theta^L - (\theta^M - \theta^L)\frac{f_1^H + f_1^M}{f_1^L} - \left(\delta^M + (\delta^M - \delta^L)\frac{f_1^H}{f_1^M}\right)\right] \\
(\theta^L, \delta^L) &: f_1^L f_1^H \left[\theta^L - (\theta^M - \theta^L)\frac{f_1^H + f_1^M}{f_1^L} - \delta^L\right]
\end{aligned}$$

Hence, by using the definition for θ_1^1 and δ_1^1 implementation of the social choice function is possible if and only if

$$f_1^L(\theta_1^M - \delta_1^M) \geq \delta_1^H - \theta_1^L \ .$$

If we introduce a third type for each player, so that $\theta_1^M < \delta_1^H$

Figure 1.9: 'Round 1', Case b



Applying Proposition 1.4, efficient bilateral trade is possible if and only if

$$f_1^L(\theta_1^H - \delta_1^L) \geq \delta_1^H - \theta_1^L \ .$$

Proof of Observation 1.7.
By Observation 1.1, we know that implementation in 'round 0' is possible if

$$f_0^L > d \ .$$

By Observation 1.9, we know that implementation in 'round 1' is possible if

$$f_1^L > \frac{\delta_1^H - \theta_1^L}{\theta_1^M - \delta_1^M} \ .$$

Making use of Definition 1.2, implementation in 'round 1' is less costly than in 'round 0', i.e., the minimal subsidy decreases, if

$$f_1^L > \frac{2f_0^L}{1 + f_0^L} \ .$$

The monotone hazard rate assumption is satisfied if

$$0 \leq \frac{f_1^H}{f_1^M} \leq \frac{f_1^H + f_1^M}{f_1^L} \ .$$

When f_1^M is sufficiently small, the monotone hazard rate assumption is violated and implementation can still be possible in 'round 1'.

Proof of Observation 1.8.
Consider the mechanism, in which the third type θ^M and δ^M for each player is added such that $\theta_1^M < \delta_1^M$. Efficient implementation is possible, whenever Corollary 1.1 applies. If an additional type is introduced for the buyer and the seller, there are 9 states of the

economy. Trade is efficient only in 5 of them.

$$
\begin{aligned}
(\theta^H, \delta^H) &: f_1^H f_1^L \left[\theta^H - \left(\delta^H + (\delta^H - \delta^M)\frac{f_1^H + f_1^M}{f_1^L}\right)\right] \\
(\theta^H, \delta^M) &: f_1^H f_1^M \left[\theta^H - \left(\delta^M + (\delta^M - \delta^L)\frac{f_1^H}{f_1^M}\right)\right] \\
(\theta^H, \delta^L) &: f_1^H f_1^H \left[\theta^H - \delta^L\right] \\
(\theta^M, \delta^L) &: f_1^M f_1^L \left[\theta^M - (\theta^H - \theta^M)\frac{f_1^H}{f_1^M} - \delta^L\right] \\
(\theta^L, \delta^L) &: f_1^L f_1^H \left[\theta^L - (\theta^M - \theta^L)\frac{f_1^H + f_1^M}{f_1^L} - \delta^L\right]
\end{aligned}
$$

Hence, implementation of the social choice function is possible if and only if

$$
f_1^L \geq \frac{\delta^H - \theta^L}{\theta^H - \delta^L} .
$$

Proof of Observation 1.9.
Consider the Mailath and Postlewaite (1990) setup for $I = \{1, 2\}, \Theta_i = \Theta = \{\theta_1^L, \theta_1^M, \theta_1^H\}$. We assume, that $0.5 \left(\theta_1^L + \theta_1^M\right) > c$. Then g can be implemented efficiently, if and only if

$$
\frac{1}{2} \sum_{i=1,2} \mathbb{E}_{(\theta)}[v(\theta_i)g(\theta)] \geq c\mathbb{E}_{(\theta)}[g(\theta)]
$$

Since trade takes place, if at least one consumer has a middle valuation, and g is either 0 or 1, the right-hand side of the equation equals

$$
\begin{aligned}
c\mathbb{E}_{(\theta)}[g(\theta)] =& c(1 - Prob(\text{both have a low valuation})) \\
&= c\left(1 - (f_1^L)^2\right) = c\left((1 + f_1^L)(1 - f_1^L)\right) = c(1 + f_1^L)(f^H + f_1^M)
\end{aligned}
$$

The left-hand side is given by:

$$
\begin{aligned}
&\frac{1}{2} \sum_{i=1,2} \mathbb{E}_{(\theta)}[v(\theta_i)g(\theta)] \\
&\qquad = \theta_1^M \left[f_1^L \left(f_1^H + f_1^M\right)\right] + \theta_1^L \left(f_1^H + f_1^M\right) \\
&\qquad = \theta_1^M f_1^L \left(f_1^H + f_1^M\right) + \theta_1^L \left(f_1^H + f_1^M\right)
\end{aligned}
$$

This yields that efficient implementation is possible, if and only if

$$
f_1^L > \frac{c - \theta_1^L}{\theta_1^M - c}
$$

2

On the durability of price and quantity mechanisms

2.1 Introduction

This chapter contributes to the theory of externalities. It looks at a model in which productive activities by firms cause environmental damages, which are harmful to consumers. Specifically, it compares different instruments a regulatory agency can use to control the damages.

There is already a rich literature to this debate. A crucial question is to derive conditions under which the regulation of externalities by means of a permit system is superior to taxation. Martin Weitzman (1974) set the starting point with his seminal paper "Prices versus Quantities". He examined the welfare implication of price and quantity instruments when avoidance costs and avoidance benefits are uncertain. Thus, there is uncertainty about the state of the economy in which the chosen policy instrument will influence the behavior of firms and consumers. Weitzman demonstrated that when the regulator's objective is to maximize the surplus of controlling the externality, the slopes of demand (that is shaped by the distribution of preferences for externality reduction) and supply (that is shaped by the distribution of costs) curves are important to determine which instrument leads to higher surplus. His basic reasoning is that the choice of instrument depends on whether the regulator wants to get the amount of externalities right or whether he wants to have certainty on the costs. Quantity regulation, via tradable permits, assures that the desired externality reduction takes place, regardless of costs. In contrast, price regulation, via corrective taxes, assures that the marginal costs of externality regulation never exceed the tax, but the overall reduction of externalities remains uncertain. In situations where it is crucial to get the amount of reduction right, a quantity instrument should be preferred over a price instrument. And a price instrument should be preferred when the marginal costs of reduction vary strongly with the amount of reduction.

However, in Weitzman's work and most of the literature that followed, incomplete, but symmetric information was assumed; i.e., the regulating agency, as well as firms and consumers are uninformed about the costs and benefits of externality regulation. This literature does not model information problems that arise if firms privately observe their avoidance costs, and consumers privately observe the harm that the environmental damage causes. But most likely firms have superior information than the regulatory agency about the avoidance costs, and also consumers will have better information about how much they benefit from reduced externalities. The literature that is based on the assumption of symmetric information does not answer the question of how a regulator would obtain the information about the costs and benefits that his intervention causes. To obtain these information he would have to communicate with consumers and firms in the economy. Both groups, on the other hand, may have an incentive to communicate strategically, so as to influence the regulator's perception of costs and benefits of externality reduction. Therefore, private information has a profound impact on the regulation of externalities.

In this chapter, I study how price and quantity instruments should be designed if information is private. For an assessment of the two instruments, I introduce an optimal unconstrained mechanism to regulate externalities. Hence, a first contribution of this chapter is to describe an optimal unconstrained mechanism when many firms can reduce the externality. I show that the optimal amount of externality reduction is determined according to the Samuelson rule (see Samuelson, 1954), so that the sum of marginal benefits of the externality reduction is equal to the marginal costs. A second contribution is to characterize price and quantity instruments under private information. I show that both instruments fail to satisfy the Samuelson rule. Specifically, the efficiency conditions, when price or quantity instruments are employed, are based on the regulator's expectations about consumers' preferences and firms' costs. Whenever these expectations do not coincide with the realized preferences and costs, price and quantity instruments fail to achieve efficiency. A third contribution of this chapter is to analyze whether an unconstrained mechanism is able to Pareto improve upon price and quantity mechanisms. I show that price and quantity mechanisms are *durable*, in the sense that there is at least one type of consumer that achieves higher utility, respectively profits under the already introduces mechanism.

The analysis of this chapter is based on the independent private values model, introduced in Chapter 1, with the following features: On the one hand, there are firms that can reduce externalities; they differ with respect to their costs. On the other hand, there are consumers that benefit from the externality reduction; they differ in their valuation for the externality reduction. Furthermore, consumers need to compensate firms for the reduction. The regulator can either use a price or a quantity instrument to control the externality, or he can use a mechanism design approach. The regulator needs to introduce the instrument before uncertainty is resolved that is before consumers' preferences and firms' costs are known. To understand the difference between the three policy instruments, I set my focus to the market of externality reduction. Specifically, I use a partial equilibrium model to derive the optimal regulation of externalities. I assume that firms can reduce externalities by using available technologies, e.g. introducing filters to cut emissions. Thus, I do not account for firms to reduce externalities by producing less output.

The regulator faces the problem of preference and cost elicitation. He has to provide incentives for a truthful revelation of privately held information. I impose that firms need to reduce externalities voluntarily. Hence, firms cannot be forced by the regulator to reduce externalities. As I excluded the possibility that firms can reduce externalities by cutting production, I assume that it is not in the regulator's interest that firms make losses when reducing externalities, and with that eventually force firms out of business. Contrary, I assume that the regulator can use his coercive power and force consumers to pay for the externality reduction, even if that makes some consumers worse off compared to the case where no externality reduction takes place. The reason for treating consumers' and firms' participation differently is first, that this looks empirically plausible that if the regulator needs to use his coercive power, he uses it to force consumers to pay a tax rather than risking that firms need to close. As consumers' participation constraints need not to be satisfied, consumers can be forced to pay for externality reduction although they might not benefit from it, and firms do receive monetary compensation. Second, this framework is most closely related to Weitzman's (1974) analysis, that does not account for changes in the demand and supply of externality reduction due to the introduction of price and quantity instruments.

The analysis proceeds as follows: First, I characterize an optimal unconstrained mechanism to control externalities as a benchmark case. I start the analysis by deriving the unconstrained mechanism and show that the surplus-maximizing externality regulation needs to fulfill the condition that the sum of consumers' marginal benefits equals the the firm's marginal costs of externality reduction. This is the classical optimality condition for a public good, derived by Samuelson (1954), when many firms provide the public good. If one thinks about CO_2-emissions as an externality, the amount of emissions experienced by one consumer is not influenced by other consumers experiencing it. Hence, such externalities have the characteristics of being non-excludable and non-rivalrous; the characteristics of a public good.

Second, I describe how the government can regulate externalities with a price instrument. Corrective pricing works such that a tax is introduced that is equal to the marginal costs of externality reduction. I derive the optimal price instrument that maximizes expected surplus and show that the Samuelson rule is no longer satisfied. Under the price regulation, a modified condition has to hold where the expected sum of consumers' marginal benefits equals the firm's marginal costs of emission reduction; i.e., the Samuelson rule is only fulfilled in expectation. The regulator fixes the subsidy before uncertainty is resolved, so that the marginal costs of emission reduction are certain, however, the amount of externality reduction is uncertain and varies with firms' real costs. The optimality condition for externality reduction, when a price mechanism is used, states that the regulator should fix the marginal costs such that they coincide with his expectation about marginal benefits. Whenever the regulators ex ante expectations on preferences and costs differ from their ex post realizations, the price instrument fails to reach efficient outcomes.

The third instrument, a regulator can use to reduce externalities, is a quantity instrument. The regulator demands a fixed amount of externality reduction. A relevant question is how the government decides to what extend a single firm should reduce externalities. I describe the quantity mechanism that maximizes expected surplus. I show that a firm's share of reduction is going to depend on its marginal costs of externality reduction. To account for firms' private information, I design the quantity instrument such that no firm has an incentive to misreport its type. The regulator fixes the amount of externality reduction, being uncertain about the preferences and costs. I demonstrate that the quantity mechanism satisfies again only a modified Samuelson rule where the sum of expected preferences times the marginal benefits of overall emission reduction equals the expected sum of marginal costs. As under the price instrument, whenever the regulator's expectations about agents' private information differ from their realization, the quantity regulation fails to reach efficient outcomes.

A mechanism design approach that can make outcomes such as the transfers and quantities depend on demand and supply for externality reduction, is more flexible than price and quantity instruments. A mechanism designer specifies the efficient externality level for every realization of costs and preferences. This observation is a challenge for externality regulation by means of price and quantity instruments because the unconstrained mechanism leads to first-best efficient outcomes. One would therefore assume that if the regulator proposes to switch from price and quantity instruments to an unconstrained mechanism, Pareto improvements will be possible.

In order to check if an optimal unconstrained mechanism to control externalities can Pareto improve upon optimal price and quantity instruments, I introduce the concept of durable contracts, in the sense of Holmström and Myerson (1983). An optimal unconstrained mechanism that does not rely on a priori assumptions on the form of regulation policy reaches surplus-maximizing outcomes. Hence, it leads to higher surplus than the before introduced price and quantity instruments. I consider whether renegotiation is possible. The outcomes of the price and quantity regulation are the default options for renegotiation. The optimally designed externality mechanism can only replace the price or quantity instrument if all agents achieve higher or equal utility, respectively profits, under the new unconstrained mechanism. If the optimal unconstrained mechanism is not unanimously accepted, then price and quantity instruments stay in place. I show that price and quantity instruments are durable; the unconstrained mechanism cannot Pareto improve upon already installed price and quantity instrument.

The remainder of the chapter is organized as follows: The next section gives a more detailed literature review. Section 2.3 specifies the economic environment. Section 2.4 introduces as a benchmark case an optimally designed unconstrained mechanism. Section 2.5 contains the description of the price and quantity mechanisms, under the assumption that there is a social planner that maximizes surplus. Section 2.6 clarifies whether a switch from a price or quantity mechanism to an optimally designed unconstrained mechanism leads to Pareto improvements. Section 2.7 relates my results to Weitzman's results. The last section contains concluding remarks.

2.2 Literature

This chapter is related to different strands of the literature. The main part of this chapter builds on the analysis of Weitzman (1974). His paper had a profound impact on the externality literature. The question if price or quantity instrument should be used has mainly been studied in the context of environmental economics (see Cropper and Oates, 1992, for a survey). However, the question, which instrument to use when agents have private information, has not been analyzed so far.

Second, this chapter draws on the literature that uses a mechanism design approach, under the constraint that firms' voluntarily reduce externalities. I do abstract from imposing the constraint that consumers' participation in the mechanism is voluntary because the mechanism design literature has shown that, under the assumption of consumers having private information about their preferences for externality reduction, there exists no mechanism that assures efficient outcomes, see e.g., Rob (1989) and Mailath and Postlewaite (1990).[1] Contrary, if the regulator can force consumers to pay for externality reduction, although they may not benefit, then efficiency can be reached even under the assumption that consumers have private information (see Arrow, 1979, and D'Aspremont and Gerard-Varet, 1979). Most of the literature on externality regulation assumes that there is one representative firm reducing the externality and the costs of reduction is commonly known. Here, I apply the framework developed in Chapter 1 as it allows me to study allocation problems where many firms that have private information

[1] Mailath and Postlewaite (1990) focus on limit results when the number of consumers goes to infinity and assume that the there is a continuum of possible preference types. Chapter 1 shows that this impossibility result holds as well under the assumption that the set of possible types is discrete.

on their costs, provide shares of a public good.

Finally, this chapter builds on Holmström and Myerson (1983) in that it provides a dynamic formalization of the question whether price and quantity instruments are durable, meaning that no other regulation can Pareto improve upon the two instruments. In particular, I want to compare the status quo results of a price or a quantity instrument with the outcomes of the unconstrained mechanism. Two recent papers have linked externality regulation and different outside options. Harstad (2012) analyzes a dynamic game in which countries make decisions on how much greenhouse gases to emit and how much to invest in technologies to reduce emissions. Further, countries can write incomplete contracts, so as that they can contract on emissions but not on technology investments. He finds that first-best implementation is possible in an incomplete contract model when renegotiation is possible. However, Harstad abstracts from a production side, private information and the possibility to opt out of climate agreements. This chapter is also related to Grüner and Koriyama (2012). They analyze whether efficient emission reduction takes place when it is necessary that each consumer is made better off relative to an outcome with majority voting about externality reduction, rather than the usual assumption that the outside option is no externality reduction. They find that efficient externality reduction is possible. My approach differs in that the threat point is either an already installed price or an already installed quantity mechanism.

2.3 The economic environment

2.3.1 Firms

There is a set of firms $J = \{1, ..., m\}$. A firm's payoff is given by

$$\pi_j(\delta_j, r_j, q_j) = r_j - \delta_j c(q_j) \ ,$$

where q_j is the externality reduction by firm j. Each firm receives a revenue r_j for reducing externalities. c is an increasing and convex cost function that satisfies $c(0) = 0$, $\lim_{q\to 0} c'(q) = 0$ and $\lim_{q\to\infty} c'(q) = \infty$, where the Inada conditions avoid corner solutions. The cost parameter δ_j is privately observed by the firm. From the perspective of all other agents it is a random variable with support $\{\delta^1, ..., \delta^r\}$, and probability distribution $(p^1, ..., p^r)$. I assume that $\delta^l < \delta^{l+1} \ \forall \ l \in \{1, ..., r-1\}$, so that firms with a lower index have a lower cost type. The distribution is assumed to be common knowledge. I denote by p^l that $\delta_j = \delta^l$ and by $P(\delta_j^l)$ the probability that $\delta_j > \delta^l$. The random variables $(\delta_j)_{j\in J}$ are assumed to be independently and identically distributed ($i.i.d.$). I write $\delta = (\delta_1, ..., \delta_m)$ for the vector of all cost parameters and δ_{-j} for the vector that lists all cost parameters except δ_j. I impose a monotone hazard rate assumption: The function $h(\delta_j) := \frac{P(\delta_j)}{p(\delta_j)}$ is assumed to be non-decreasing.

2.3.2 Consumers

The set of consumers is denoted by $I = \{1, ..., n\}$. The preferences of consumer i are given by the utility function:

$$u_i(\theta_i, Q, t_i) = \theta_i b(Q) - t_i \ ,$$

where $Q = \sum_{j=1}^{m} q_j \in \mathbb{R}_+$ denotes the total externality reduction, t_i is the monetary payment of consumer i to the firms and θ_i is consumer i's preference parameter that belongs to a finite ordered set $\Theta = \{\theta^0, \theta^1, ..., \theta^s\}$. I assume that $\theta^l < \theta^{l+1} \ \forall \ l \in \{0, ..., s-1\}$, etc. The consumer privately observes θ_i. From the perspective of all other agents it is a random variable with support Θ and probability distribution $(f^0, ..., f^s)$. I denote by f^l that $\theta_i = \theta^l$ and by $F(\theta_i^l)$ the probability that $\theta_i > \theta^l$. This distribution is assumed to be common knowledge. The random variables $(\theta_i)_{i \in I}$ are assumed to be *i.i.d.*. I write $\theta = (\theta_1, ..., \theta_n)$ for a vector of all preference parameters and θ_{-i} for a vector that lists all preference parameters except θ_i. I impose a monotone hazard rate assumption, so that the function $g(\theta_i) = \frac{1-F(\theta_i)}{f(\theta_i)}$ is assumed to be non-increasing.

Firms' cost parameters and consumers' preference parameters are assumed to be stochastically independent. I refer to a vector (θ, δ) that lists all preference and technology parameters as state of the economy. In the following, whenever I use the expectations operator, this indicates that expectations are taken with the joint probability distribution of δ and θ.

2.3.3 Mechanism

A mechanism design approach is used to characterize the reduction level and the pricing of externalities. I restrict myself to direct mechanisms, by appealing to the revelation principle (see Myerson, 1985), so that a truthful revelation of costs by firms and of preferences by consumers is a Bayes-Nash equilibrium.

A direct mechanism is a collection of functions $(q_j, r_j)_{j=1}^{m}$ and $(Q, t_i)_{i=1}^{n}$, where $q_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}_+$ characterizes j's externality reduction as a function of the vector of agents' reports of their types to the mechanism designer. Analogously, $r_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}$ determines j's monetary revenue as a function of the vector of agents' reports. And the function $t_i : \Delta^n \times \Delta^m \mapsto \mathbb{R}$ specifies i's payment as a function of the vector of consumers' preference parameters and firms' cost parameters; $Q = \sum_{j=1}^{m} q_j$ specifies the level of the overall externality reduction.

Incentive compatibility constraints. Truth-telling of firm j is a best response if, for all l and all k,

$$R_j(\delta^l) - \delta^l C_j(\delta^l) \geq R_j(\delta^k) - \delta^l C_j(\delta^k), \qquad (IC_F)$$

where $R_j(\delta^l) \equiv \mathbb{E}_{(\theta,\delta_{-j})}\left[r_j(\theta,\delta_{-j},\delta_j)|\delta_j = \delta^l\right]$ is the expected revenue for firm j and $C_j(\delta^l) \equiv \mathbb{E}_{(\theta,\delta_{-j})}[c_j(q_j(\theta,\delta_{-j},\delta_j))|\delta_j = \delta^l]$ are the expected production costs, in case of reporting a cost parameter δ^l, given that all other firms and consumers reveal their private characteristics to the mechanism designer. Due to the assumption that firms' cost parameters and consumers' preference parameters are stochastically independent, all firms have the same belief about δ_{-j} and θ. Thus, I can examine j's expected transfers and production level as a function that only depends on j's announcement.

In the same way, truth-telling of consumer i is a best response if, for all l and k,

$$\theta^l B_i(\theta^l) - T_i(\theta^l) \geq \theta^l B_i(\theta^k) - T_i(\theta^k), \tag{IC_C}$$

where $B_i(\theta^l) \equiv \mathbb{E}_{(\theta_{-i},\delta)}\left[b_i(Q(\theta_{-i},\theta_i,\delta))|\theta_i = \theta^l\right]$ are the expected benefits of externality reduction, when consumer i announces θ^l and $T_i(\theta^l) \equiv \mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\theta_{-i},\theta_i,\delta)|\theta_i = \theta^l\right]$ are the expected payments, in case of reporting a preference parameter of θ^l.

Participation constraints. A mechanism also has to satisfy participation constraints, which ensure that firm profits from the externality reduction are bigger than a minimal profit requirement.

The interim participation constraints for firms are: For all j, and all l,

$$R_j(\delta^l) - \delta^l C_j(\delta^l) \geq \underline{\pi}_j, \tag{PC_F}$$

where $\underline{\pi}_j$ denotes a lower bound for firm j's profit in order to make it participate voluntarily in the mechanism. These constraints ensure that after all firms have discovered their own cost parameter, no firm is worse off relative to a status quo situation in which no externalities needed to be reduced and profits were $\underline{\pi}_j$. An alternative interpretation is that firms have veto rights that protect them from receiving revenues that do not cover the costs of externality reduction. Consequently, a deviation from the status quo calls for an unanimous agreement of all firms to reduce the externality. Additionally, these constraints ensure that the number of firms on the interim stage is exogenously given because no firm is forced out of business as it needs to reduce externalities.

For the consumers, however, I require no participation constraints. I assume that the regulator can rely on his coercive power when financing the reduction of externalities. I impose no participation constraints, as it has been shown by Mailath and Postlewaite (1990) that with many consumers no externality reduction will take place if consumers' participation is voluntary. The reason is that, with many consumers, a single consumer's influence on the externality reduction is close to zero. Due to voluntary participation constraints, consumers can drive their transfers down to zero. Hence, insufficient transfers are collected to finance the externality reduction.

Budget constraint. I assume that the direct mechanism has to satisfy a budget constraint. It requires that the consumers' expected transfers are larger than the firms' expected revenues,

$$\mathbb{E}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] \geq \mathbb{E}\left[\sum_{j=1}^{m} r_j(\theta,\delta)\right]. \qquad (BC)$$

2.4 Optimal Mechanism Design

Before I introduce price and quantity instruments to regulate externalities, an unconstrained mechanism to externality regulation will be defined as a benchmark. As a first step, I provide a necessary and sufficient condition for a mechanism to be incentive-compatible and budgetary feasible and in addition satisfies firms' participation constraints.

Proposition 2.1. *Let Q be a reduction level with the property that for all i, and all l, the monotonicity constraints $B_i(\theta^l) \geq B_i(\theta^{l-1})$ are satisfied and q_j be a production rule with the property that for all j, and all k, the monotonicity constraints $C_j(\delta^{k-1}) \geq C_j(\delta^k)$ are satisfied. There exists a mechanism that is incentive compatible and that fulfills (PC_F) if and only if*

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b(Q(\theta,\delta))\right] \geq \mathbb{E}\left[\sum_{j=1}^{m}\left(\delta_j + \frac{P(\delta_j)}{p(\delta_j)}\right) c(q_j(\theta,\delta))\right] + \sum_{j=1}^{m} \underline{\pi}_j .$$

A proof can be found in the Appendix. The term $\left(\delta_j + \frac{P(\delta_j)}{p(\delta_j)}\right) c(q_j(\theta,\delta))$ is known as firm j's virtual costs for the externality reduction. Virtual costs can be interpreted as the minimal amount that type δ_j of firm j needs to receive for the externality reduction when incentive compatibility and participation constraints need to be satisfied. In a setting with complete information, the minimal amount would be the true costs δ_j. With private information on costs this expression is increased by the hazard rate $\left(\frac{P(\delta_j)}{p(\delta_j)}\right) c(q_j(\theta,\delta))$, which can thus be interpreted as an information rent. That consumers do not achieve information rents relies on the fact that their participation constraints do not need to be satisfied; i.e., the regulator is able to force consumers to finance of externality reduction. Proposition 2.1 says that Pareto-efficient externality reduction is possible if and only if the expected benefits of externality reduction exceed the expected virtual costs. Proposition 2.1 applies to any decision rule Q such that B_i is increasing for all i, and q_j such that C_j is decreasing for all j.

There is one mechanism that is of particular interest, namely the mechanism that maximizes expected surplus[2]

$$S(\theta,\delta) = \mathbb{E}\left[\sum_{i=1}^{n}\left(\theta_i b(Q(\theta,\delta))\right) - \sum_{j=1}^{m}\left(\delta_j c(q_j(\theta,\delta))\right)\right] .$$

The following Proposition introduces the optimal unconstrained mechanism given that a mechanism design approach was chosen to regulate the externality.

Proposition 2.2. *The optimal unconstrained mechanism that maximizes $S(\theta,\delta)$, in the following denoted by $\{(q_j^*, r_j^*)_{j=1}^m, (Q^*, t_i^*)_{i=1}^n\}$, subject to the constraints in (IC_F), (IC_C), (PC_F) and (BC), has the following properties:*

i) Consumer surplus under the optimal unconstrained mechanism is equal to

$$U(\theta,\delta) = \mathbb{E}\left[\sum_{i=1}^{n}\left(\theta_i b(Q^*(\theta,\delta))\right) - \sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j^*(\theta,\delta))\right] - \sum_{j=1}^{m}\underline{\pi}_j .$$

ii) Expected firm profits are

$$\Pi(\theta,\delta) = \mathbb{E}\left[\sum_{j=1}^{m} h(\delta_j)c(q_j^*(\theta,\delta))\right] + \sum_{j=1}^{m}\underline{\pi}_j .$$

iii) Output is undistorted. For every (θ,δ), the reduction level $q_j^*(\theta,\delta)$, for all firms j, satisfies

$$\sum_{i=1}^{n}\theta_i b'\left(\sum_{j=1}^{m} q_j^*(\theta,\delta)\right) = \delta_j c'(q_j^*(\theta,\delta)) .$$

The mechanism designer can use his coercive power and force consumers to pay for the externality reduction. Thus, the efficient regulation comes at costs for consumers. Consumers' transfers do not only need to cover the costs of reduction but as well firms' information rents, due to the fact that firms have private information about their costs and that their profits need to be at least as high as the minimal profit requirement $\underline{\pi}_j$. The optimal unconstrained mechanism violates consumers' participation constraints when the virtual costs of externality reduction exceed consumers virtual valuation of the externality reduction.[3] For consumers that benefit only little from externality reduction, the participation constraints are always violated.

The Proposition demonstrates that first-best efficient externality reduction can be attained if consumers and firms have private information on their characteristics. Exter-

[2]For a generalization of the surplus function, see Chapter 3.

[3]Similarly to firms' virtual costs, consumer i's virtual valuation for externality reduction is given by $\theta_i - \frac{1-F(\theta_i)}{f(\theta_i)}$. This term can be interpreted as the maximal amount that consumer i with type θ_i will pay when incentive compatibility constraints and participation constraints need to be respected.

nalities should be reduced up to the point where the sum of marginal benefits equals the marginal costs of externality reduction; this result has been established by Samuelson (1954) for a model of public good provision, with no private information on consumers' preferences and firms' costs. A model of externality reduction can be translated into a model of public good provision if one thinks about non-depletable externalities.[4] Proposition 2.2 iii) is the classic optimality condition for public good provision; known as Samuelson rule, given that many firms provide parts of the public good.

2.5 Price and quantity mechanisms under asymmetric information

The regulating agency of the preceding section could make the externality dependent on changes in costs and preferences. For every state of the economy (θ, δ), the efficient externality reduction takes place and the Samuelson rule is satisfied. However, in real-world applications, externality regulation is not based on planning that conditions the reduction on the future state of the economy. Rather, price and quantity instruments are introduced before the uncertainty is resolved. Then, it becomes a relevant question to ask whether asymmetric information has an impact on the regulating choice between price and quantity instruments.

The discrepancy in efficiency between prices and quantities was firstly analyzed by Weitzman (1974). In his seminal paper, "Prices versus Quantities", he shows that under incomplete information both instruments differ with respect to efficiency. If a price instrument is chosen to regulate externalities, the regulator sets a price that a firm gets per unit of reduction. Weitzman assumes that the regulator sets this price based on his expectation about the future state of the economy. Further, he assumes that the regulator can commit to not change the price anymore. As the regulator fixes the price, marginal costs of externality reduction are certain but uncertainty remains about the marginal benefits. If instead the regulator fixes the quantity, then the marginal benefits will be certain but uncertainty will remain about the marginal costs. Weitzman derives a comparative advantage measure to show the efficiency gain of using one instrument over the other. In section 2.7, I show how Weitzman's setup relates to the one of this chapter.

As Weitzman, I am interested in a regulator that wants to maximize total surplus when marginal costs and marginal benefits are uncertain at the moment the regulator needs to decide whether to employ a price instrument or a quantity instrument to regulate the externality. In contrast to Weitzman, I do assume that information is asymmetric and agents may have an incentive to communicate strategically. The regulating agency

[4] An externality is said to be non-depletable if one consumer's benefit of the externality reduction does not diminish another consumer's ability to benefit from the reduction; see e.g. Chapter 11 Mas-Colell, Whinston, and Greene (1995).

can either set a price that firms receive per amount of externality reduction or set the quantity by fixing the overall amount of externality reduction. Price and quantity instruments are designed optimally under the assumption that the regulator remains uninformed about the current state of preferences and costs.

The main insight of the analysis below is that optimally designed price and quantity mechanisms fail to achieve first-best efficient outcomes; the Samuelson rule is no longer satisfied. I will show that the Samuelson rule under price and quantity instruments is only fulfilled in expectation.

2.5.1 Price Regulation

One way for the regulator to control externalities is to impose a subsidy on the externality reducing activity. This mechanism is known as Pigouvian taxation, see Pigou (1954). I proceed in two steps: First, I take the marginal costs of externality reduction as given and solve for the price instrument reaching that target, before I identify the optimal price for externality reduction in a second step.

Reaching a given target. When a subsidy s is announced, every firm will decide how much externality reduction to contribute. Every firm is maximizing its payoffs that is

$$\pi_j(\delta_j, s) = s q_j(\delta_j, s) - \delta_j c(q_j(\delta_j, s)) \ .$$

Every firm j therefore chooses $q_j(\delta_j, s)$ according to the first order condition

$$s - \delta_j c'(q_j(\delta_j, s)) \leq 0 \ ,$$

with equality for $q_j(\delta_j, s) > 0$. These conditions determine the optimal amount $q_j(\delta_j, s)$ every firm j reduces, which depends solely on the announced subsidy and firm j's cost parameter δ_j. It is firm j's reaction function to a subsidy s. The other firms' cost parameters are not relevant for firm j's externality reduction q_j, which is different under the unconstrained mechanism. Note that $q_j(\delta_j, s)$ is maximizing firm j's profit. Whenever firm j deviates to $q_j(\delta_j', s)$, profits are declining. Thus, firm j's incentive compatibility constraints in (IC_F) are satisfied.

Further, the price mechanism assures voluntary participation because firms would decide to not reduce any part of the externality reduction if this had come along with losses. The resource costs under a price instrument are the subsidies paid times the expected sum of externality reduction and firms' minimal resource requirement

$$R(\delta, s) = s\mathbb{E}\left[\sum_{j=1}^{m} q_j(\delta_j, s)\right] + \sum_{j=1}^{m} \underline{\pi}_j = s\mathbb{E}[Q(\delta, s)] + \sum_{j=1}^{m} \underline{\pi}_j \ .$$

The optimal target. The subsidy s is determined by the regulator before consumers learn their preferences and firms learn their costs. As under the unconstrained mechanism, the regulator has no outside funding option, so that expected consumer transfers need to cover the expected resource costs $\mathbb{E}\left[\sum_{i=1}^{n} t_i\right] \geq R(\delta, s)$. A given subsidy therefore induces a level of total surplus that equals

$$S(\theta, \delta, s) = \mathbb{E}\left[\sum_{i=1}^{n} \theta_i b(Q(\delta, s)) - \sum_{j=1}^{m} \delta_j c(q_j(\delta_j, s))\right] .$$

The following Proposition characterizes the surplus-maximizing price mechanism.

Proposition 2.3. *The surplus maximizing price mechanism has the following properties:*

i) Expected consumer surplus is equal to

$$U(\theta, \delta, s^*) = \mathbb{E}\left[\sum_{i=1}^{n} \theta_i b(Q(\delta, s^*))\right] - R(\delta, s^*) .$$

ii) Expected firm profits are

$$\Pi(\delta, s^*) = R(\delta, s^*) - \mathbb{E}\left[\sum_{j=1}^{m} \delta_j c(q_j(\delta_j, s^*))\right] .$$

iii) Externality reduction fails to satisfy the Samuelson rule. The subsidy s^* satisfies the first order condition

$$s^* = \mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'(Q(\delta, s^*))\right] .$$

Expected consumer surplus does not only depend on preferences and costs, as in Proposition 2.2. When a price mechanism is used to regulate emissions, the introduced subsidy influences expected consumer surplus and expected firm profits.

The Proposition shows that the Samuelson rule is violated. Here the Samuelson rule is only satisfied in expectation, as the marginal costs of externality reduction s need to equal the expected sum of marginal benefits. When the marginal benefits differ from the ex ante expectations, first-best efficient externality reduction cannot be achieved with a price instrument. For $\sum_{i=1}^{n} \theta_i b'(Q(\theta, \delta)) > \mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'(Q(\delta, s^*))\right]$, the subsidy s^* is lower than the subsidy that leads to ex post efficiency. Hence, ex post, there is not enough reduction of externalities.

This expected Samuelson rule is well suited to describe how Weitzman's (1974) analysis is related to my setup. As the regulator fixes the price, the marginal costs of externality reduction are certain. Uncertainty remains about the amount of externality reduction. If demand (that is shaped by consumers' preferences for emission reduction) and supply (that is shaped by firms' costs) differ from the expected demand and supply, only the amount of emission reduction can adapt to these changes. Contrary to the unconstrained mechanism, the price cannot change, so that a surplus loss occurs. How big the surplus loss under a price instrument is, depends on how much the ex ante expectations on costs and preferences differ from the ex post realization. The more these values differ, the more severe is the surplus loss of a price instrument in comparison to the unconstrained mechanism, and the higher is the misspecification of externality reduction.[5]

2.5.2 Quantity Regulation

Beside a price mechanism, the regulator can as well control externalities with a quantity mechanism. That is, the regulator decides ex ante on the total amount of externality reduction Q and the production sector needs to provide this reduction. The quantity mechanism should not be understood such that the regulator asks firms about their types and decides by how much the externalities should be reduced by every single firm. The timing is such that the regulator has expectations about firms' cost parameters and consumers' preference parameters and decides on the ex ante stage how much externalities should be reduced. The total amount of externality reduction is then subdivided into parts and the fraction each firm reduces depends on its announced cost parameter. In the following I seek to characterize the mechanism with this property. Again, I proceed in two steps. First, I take the quantity goal as given, and solve for the quantity mechanism reaching that target, before I identify the optimal externality reduction in a second step.

Reaching a given target. I am interested in reaching the externality reduction that maximizes expected surplus. Beside consumers' and firms' incentive compatibility constraints, firms' participation constraints and the budget constraint, I impose a quantity constraint, i.e., for all δ,

$$\sum_{j=1}^{m} q_j(\delta, \hat{Q}) = \hat{Q} \ , \tag{QU}$$

so that the externality reduction of all firms need to add up to the set target $\hat{Q}$.

The following Proposition characterizes the solution to this problem.

Proposition 2.4. *Suppose that a monotone hazard rate assumption holds ($h(\delta_j)$ is a non-decreasing function). The surplus-maximizing quantity mechanism that satisfies the constraints in (IC_F), (PC_F), (IC_C), (BC) and (QU), has the following properties:*

[5] For a detailed analysis, see Chapter 3.

i) The expected consumer surplus under the optimal quantity mechanism is

$$U(\theta, \delta, \hat{Q}^*) = \mathbb{E}\left[\sum_{i=1}^{n}\Big(\theta_i b(\hat{Q}^*)\Big) - \sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\delta, \hat{Q}^*))\right] - \sum_{j=1}^{m}\underline{\pi}_j \; .$$

ii) Expected firm profits are

$$\Pi(\delta, \hat{Q}^*) = \mathbb{E}\left[\sum_{j=1}^{m}\Big(h(\delta_j)c_j(q_j(\delta, \hat{Q}^*))\Big)\right] + \sum_{j=1}^{m}\underline{\pi}_j \; .$$

Proposition 2.4 shows that the ex ante set amount of externality reduction influences expected consumer surplus and expected firm profits, contrary to the unconstrained mechanism, derived in Proposition 2.2, where outcomes depended only on preferences and costs for emission reduction. Under the quantity mechanism, benefits of emission reduction are certain, but consumer surplus and firm profits will vary with preferences and costs, respectively.

The optimal target. I assume that the externality target is formulated before the consumers' preferences and firms' costs are drawn. The following Proposition shows what the quantity mechanism looks like.

Proposition 2.5. *The quantity mechanism that maximizes $S(\theta, \delta, \hat{Q})$ subject to the constraints in (IC_F), (PC_F), (IC_C), (BC) and (QU) has the following property: Externality regulation fails to satisfy the Samuelson rule. The externality reduction $\hat{Q}^*$ satisfies the first order condition*

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i\right] b'(\hat{Q}^*) = \mathbb{E}\left[\sum_{j=1}^{m}\delta_j c'(q_j(\delta, \hat{Q}^*))\right] ,$$

where $c'(q_j(\delta, \hat{Q}^*)) = \frac{\partial c(q_j(\delta, \hat{Q}^*))}{\partial q_j(\delta, \hat{Q}^*)} \frac{\partial q_j(\delta, \hat{Q}^*)}{\partial \hat{Q}^*}$.

An optimal choice of externality reduction, under the quantity mechanism, requires again that a Samuelson rule holds in expectation, i.e., that the sum of marginal benefits times the expected preferences is equal to the sum of expected marginal costs of externality reduction. As $\hat{Q}^*$ is fixed, marginal benefits are certain but uncertainty remains about the marginal costs and consumers preferences. The price at which the reduction is achieved can vary.

Again, the condition in Proposition 2.5 is well suited to describe the relation to Weitzman's (1974) paper to my setup. The quantity restriction has the effect that the actual level of externality reduction is not sensitive to changes in θ and δ as it would be required by optimality (see Proposition 2.2 iii)). When the regulator's ex ante expectation on preferences and costs differ from the ex post realizations, the realized demand and supply will differ from the ex ante expected ones. A sensitive adaption of the quantity restriction would therefore be optimal. However, as the quantity is fixed, the quantity mechanism fails to achieve ex post efficient outcomes. How big the loss in efficiency is, when using a quantity mechanism instead of the optimal unconstrained mechanism, depends on how much the optimal externality level varies with the regulator's expectation about θ and δ.

2.6 On the durability of price and quantity mechanisms

The results of the previous sections have shown the following: An optimally designed unconstrained mechanism can achieve ex post efficiency; for all possible states of the economy, the mechanism specified the optimal amount of externality reduction and the respective transfers. On the other hand, price and quantity mechanisms were less flexible. Once the regulator fixed the subsidy only the amount of externality reduction could adapt to changes in the economy, and if the regulator fixed the quantity, only the costs could adapt to changes. Therefore, whenever the regulator's expectation on preferences and costs differs from the realized values, ex post inefficient externality regulation occurs. According to the Coase theorem, see Coase (1960), ex post inefficient price and quantity mechanisms should not be observed when information is complete and when bargaining is costless. All agents in the economy should be willing to switch to an unconstrained mechanism that leads to ex post efficiency. However, Holmström and Myerson (1983) have shown that this concept does not necessarily translate to an environments with private information. In the following I am interested in the question whether an unconstrained mechanism can Pareto improve upon price and quantity mechanisms.

2.6.1 Durability

In settings of complete information, the notion of Pareto efficiency is clearly defined. If all consumers and firms consider the Pareto superior unconstrained mechanism to regulate externalities to a status quo with price and quantity instruments, then an unanimous agreement to change from the status quo to the unconstrained mechanism should be possible. Holmström and Myerson (1983) extend that concept to settings with private information. They consider the following game where all agents vote simultaneously for either the status quo or a given alternative; in the case of this chapter the unconstrained mechanism. After the vote, the unconstrained mechanism will only be implemented if all consumers and firms vote for it.

In the definition of Holmström and Myerson (1983), a price or quantity mechanism is durable if and only if, when it is the status quo, for the unconstrained externality regulating mechanism being the alternative, it is an equilibrium of the voting game that at least one type of consumer or one type of firm surely votes for the status quo, for all possible states of the economy.

I will distinguish between two concepts of efficiency: ex post when all private information is common knowledge and interim, when consumers and firms have learned their own types, respectively.

2.6.2 The main result

To check whether price and quantity mechanisms are durable, I consider the following setup: The regulating agency decides to either use a price or a quantity mechanism to reduce externalities. Then, consumers and firms learn their types. To see whether the unconstrained mechanism is able to interim Pareto improve upon the status quo, the regulating agency asks all consumers and firms to vote simultaneously. I use the initial price respectively quantity mechanism as the default option, so that consumers as well as firms can reject the new unconstrained mechanism and with that the old regulation persists. The new mechanism can therefore be interpreted as a take-it-or-leave-it offer. This means that the new mechanism replaces the initial regulation if all agents are at least as well off as under the initial mechanism but one agent is strictly better off. To see whether the unconstrained mechanism interim Pareto improves upon price or quantity mechanism, the regulating agency lets all consumers and all firms vote when they learned their own types but remain uninformed about the outcomes of price and quantity mechanisms. If a single agent votes for the status quo, then the price or the quantity instrument stays in place. In order to see whether ex post Pareto improvements are possible, the regulating agency lets all agents vote when all information is common knowledge. The following Proposition summarizes the outcomes.

Proposition 2.6.

i) *The unconstrained mechanism can neither ex post Pareto improve a quantity mechanism nor a price mechanism.*

ii) *The unconstrained mechanism can neither ex interim Pareto improve a quantity mechanism nor a price mechanism.*

The Proposition shows that although the unconstrained mechanism is efficient, it cannot replace price or quantity mechanisms. Price and quantity mechanisms are hence *durable.* To provide an intuitive understanding of this result, I discuss the main steps of the analysis.

Ex post efficiency. For an arbitrary consumer i, I compare the ex post utility when an unconstrained mechanism is used to the ex post utility when a price mechanism or a quantity mechanism is used. Similarly, for any firm j, I compare the ex post profit of the unconstrained mechanism with the ex profit when a price or a quantity mechanism was introduced. If one agent is worse off under the unconstrained mechanism, then the price mechanism, respectively the quantity mechanism, is durable. Comparing Proposition 2.2 with Propositions 2.3 and 2.5, respectively, the total amount of emission reduction under the unconstrained mechanism does not coincide with the total amount of emission reduction when price and quantity mechanisms are used.

For the case where the unconstrained mechanism leads to more emission reduction than the price or quantity mechanism, a consumer with low preferences for externality reduction does not benefit from surplus-maximizing externality reduction. As price and quantity mechanism satisfy production efficiency, the unconstrained mechanism cannot achieve more emission reduction at lower production costs. If firms reduce more emissions under the unconstrained mechanism, they need to receive higher transfers than under the status-quo mechanism. In order to assure that the optimal unconstrained mechanism assures budget balance, consumers' transfers need to be higher under the unconstrained mechanism than under the status quo. The disutility of higher transfers therefore makes a consumer with low preferences for the externality reduction veto the unconstrained mechanism.

Consider next the case where the optimal unconstrained mechanism leads to less externality reduction than the price mechanism, as well as the quantity mechanism. This implies as well that there is at least one firm that needs to reduce less emissions under the unconstrained mechanism compared to the status quo. But firms' profits are increasing in the amount of reduction, so that it vetoes the unconstrained mechanism. This firm prefers a larger fraction of externality reduction of a second-best price or quantity mechanism over a smaller fraction of the unconstrained mechanism that reaches first-best efficiency.

Consumers' realized preferences and firms' realized technologies do not affect the design of the unconstrained mechanism. This may be questioned on the following ground: The incentive compatibility constraints serve to ensure that consumers and firms announce their types truthfully. If the outcome of the price and quantity mechanisms are known, then the regulator knows all private information. Accordingly, a mechanism might be designed that makes use of the public information. However, if consumer and firms predict this behavior of the regulator, they might refrain from communicating their private information truthfully in the first place. Holmström and Myerson (1983) note that if agents are selecting a decision rule after they have learned their private information, then interim efficiency is a reasonable requirement.

Interim efficiency. To analyze whether an unconstrained mechanism can Pareto improve upon price and quantity mechanism on the interim stage, the following sequential structure is imposed: Consumers and firms have learned their own types but remain uncertain about the other agents' types. After having learned their own type agents need to vote for either the status quo, with the price or the quantity mechanism in place, or the unconstrained mechanism. Proposition 2.6 ii) shows that there is at least one consumer or firm that is better off under the status quo, so that price and quantity mechanisms stay in place.

The main argument. To understand why Pareto improvements are not possible, I discuss the main steps of the formal analysis. The expected profit of firm j, from an interim perspective are derived as follows: when interim participation constraints are respected, they get an information rent, which equals $\mathbb{E}\left[\frac{P(\delta_j)}{p(\delta_j)}c(q_j)\right]$. Hence, a firm's expected profit is increasing in the share of externality reduction q. That implies, in particular, that the willingness of firm j to switch from the quantity mechanism to the unconstrained mechanism depends on its cost parameter. If a firm has low costs of emission reduction, it expects more reduction under the unconstrained mechanism than under the quantity mechanism and is hence voting for the unconstrained mechanism, ceteris paribus. Contrary, a firm with high costs expects less reduction under the unconstrained mechanism and therefore votes for the quantity mechanism. In order to make firms with high costs not veto the unconstrained mechanism, transfers needed to increase under the unconstrained mechanism. This is, however, not possible because a consumer that benefits only little from externality reduction vetoes against mechanisms where consumers' transfers are increased. These derivations imply that there is at least one agent that is better off under the already installed quantity mechanism. Hence, the quantity mechanism is *durable*.[6]

The price mechanism is as well *durable*. On the interim stage, a firm j knows its profits under the price mechanism. That is, only if all firms expect that they make at least as high profits under the unconstrained mechanism, a switch from the price mechanism to the unconstrained mechanism is possible. As production efficiency is given under the price mechanism, the emission reduction cannot be achieved at lower costs. Profits under the unconstrained mechanism can therefore only be higher if either less emission reduction takes place and the expected revenues do not decline, or if the emission reduction level stays the same but the expected revenues increases. In the latter case, this implies as well that consumers need to pay higher transfers in order to finance the revenue requirement, which implies as well an increase of consumers' expected transfers. Knowing that firms would only vote for the unconstrained mechanism if they expect at least as high profits as under the price mechanism, a consumer with low preferences for emission reduction, votes against the unconstrained mechanism. For the second case, where the revenue requirement stays the same but less emission reduction is reduced, all consumers that have high preferences for emission reduction would veto the unconstrained mechanism.

[6]The concept of *durable* mechanisms is restricted by the assumption that only one alternative will be considered to the status quo, see e.g. Crawford (1985). Nevertheless, the unconstrained mechanism always leads to ex post efficient results and is therefore a good focal point when agents can negotiate efficiently.

2.7 Extension

In his seminal paper "Prices versus Quantities", Weitzman (1974) explored the question of whether it is better to regulate externalities by the means of a price or a quantity instrument under the assumption that redefining instruments is not possible. He showed that uncertainty concerning the marginal benefits and marginal costs of externality reduction affects the choice between the two regulatory instruments. In this section I will discuss how "Prices versus Quantities" by Weitzman (1974) relates to this chapter's setup.

In Section 2.5 I have introduced a model of asymmetric and incomplete information. The setup of Weitzman (1974), however, differs from mine in that he does not consider the agents to have private information. In his paper, uncertainty stems from incomplete information on marginal cost and benefit curves. In this chapter's framework, the marginal benefit curve is derived from consumers valuation for emission reduction and the marginal cost curve is determined by the production sector that has to reduce emissions. As Weitzman (1974), I assume that policy is defined on the ex ante stage, e.g. before agents learn their types and only the ex ante likelihoods of θ and δ are known by the regulating agency. A consumer's goal is to maximize expected utility, with expectation about future types. Respectively, a firm's objective is to maximize payoffs.

In the following, I will look at an example and assume that there is uncertainty on the ex ante stage about the parameters that shape the marginal cost and benefit curves.[7] As Weitzman (1974) I assume that β is a random variable whose distribution is known to the regulating agency but whose realization is not observed at the time price and quantity mechanisms are announced. Instruments are defined under the assumption that zero possibility is assigned to the event that the parameter β is changing. Thereafter, I check whether a price or a quantity instrument leads to higher surplus. Additionally to Weitzman (1974), I depict how surplus is changing if β differs from the regulator's ex ante expectation.

Example: *There is one firm to which is delegated the reduction of CO_2-emissions. The emission reduction q_j leads to costs $\delta_j c(q_j)$, where δ_j is a cost characteristic of the firm that belongs to the set $\Delta = \{\delta^L, \delta^H\}$ and c is a quadratic cost function, $c(q_j) = \beta \frac{1}{2} q_j^2$. The probability that the firm has low costs is denoted by p^L and accordingly $p^H = 1 - p^L$ denotes that the firm has high costs. n consumers benefit form the emission reduction. The benefit function $b(Q) = \phi Q^{1/2}$ is assumed to be concave, where $Q = \sum_{j=1}^{m} q_j$ denotes the amount of emission reduction and ϕ is a random variable. The regulating agency knows the distribution. θ_i is consumer i's preference parameter and belongs to the finite ordered set $\Theta = \{\theta^L, \theta^H\}$, where f^L denotes the probability that consumers have low preferences and $f^H = 1 - f^L$ denotes the probability that consumer have high preferences for the emission*

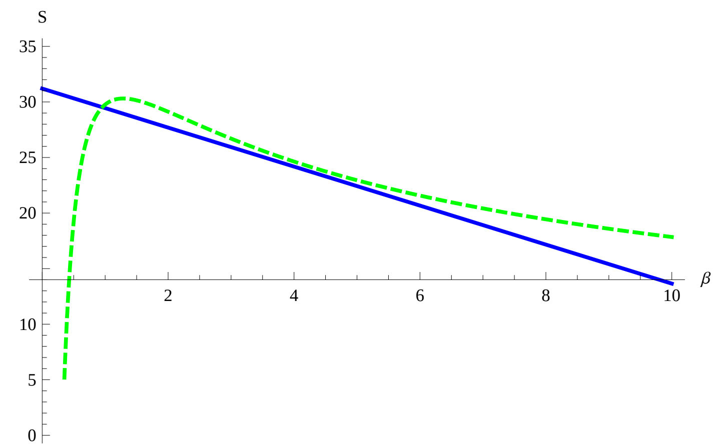
[7] Functional forms where chosen such that the Weitzman (1974) assumptions of $b'' < 0$ and $c''(q_j) > 0$ are met.

reduction.

The following Observation highlights that the cost parameter β plays a crucial role when deciding which instrument leads to higher surplus.[8]

Observation 2.1. *Consider the Example. Suppose that $S(\theta, \delta, \hat{Q}, \beta)$ and $S(\theta, \delta, s, \beta)$ are continuous functions on $[0, 10]$. Then, there exists a range $\beta \in [\hat{\beta}, 10]$ such that a price mechanism leads to higher expected surplus than the quantity mechanism. For $\beta \in [0, \hat{\beta}]$ a quantity mechanism leads to higher surplus than the price mechanism.*

Figure 2.1: Prices versus quantities



The parameter choices are $[n(f_L\theta_L + f_H\theta_H)] = 30$, $p_L = p^H = 0.5$, $\delta^L = 2$,$\delta_H = 4$, $\phi = 1$, $\hat{\beta} = 5$. The straight blue line depicts the expected total surplus function under a quantity mechanism. The dashed green curve shows the expected total surplus under a price mechanism.

Figure 2.1 illustrates the implication of a change in β graphically. If β coincides with the regulator's expected $\beta^* = 5$, the price instrument leads to the same surplus than the quantity mechanism. The same holds true for a threshold-value of $\beta = 1$. However, for $\beta > 1$ a price instrument leads to higher surplus. In comparison to Weitzman (1974), I do not just consider a deviation from β to some alternative β' but rather compare the competitive surplus advantage of one instrument over the other for the range of $\beta \in [0, 10]$. That one instrument is not outperforming the other, independently of the specification, can be explained by the fact that the surplus function under a price mechanism[9]

$$S(\theta, \delta, s, \beta) = [n(f_L\theta_L + f_H\theta_H)]^{4/3} \left[p^L \left(\frac{s}{\beta\delta_L} \right)^{\frac{1}{2}} + p^H \left(\frac{s}{\beta\delta_H} \right)^{\frac{1}{2}} \right] - \frac{1}{2}s^2 \left(\frac{p^L}{\beta\delta^L} + \frac{p^H}{\beta\delta^H} \right) ,$$

[8] I have assumed that firms are able to reduce emissions and therefore are the ones that respond to regulation. If instead I had allowed for consumers to be able to reduce emissions, e.g. via better heating technologies, the parameter ϕ would become decisive.

[9] See Chapter 3, for a derivation.

is not monotonically increasing in β. The higher β, the less emissions a single firm is reducing ($\frac{\partial q_j}{\partial \beta} < 0$), so that the expected costs of emission reduction are decreasing in β. But at the same time, the benefits are decreasing as the total amount of emission reduction decreases. If β is marginally decreased from β^*, the lower cost outweigh the lower benefits of emission reduction. If β is decreasing further, the loss in benefits is higher than the reduced costs, so that surplus is decreasing. For $\beta > \hat{\beta}$, costs of emission reduction are getting bigger, and benefits in a different emission level are even smaller, so that surplus is decreasing the bigger the difference between β and $\hat{\beta}$.

Under a quantity mechanism, however, the expected provision level is certain and with that the benefits of emission reduction

$$S(\theta, \delta, \hat{Q}, \beta) = [n(f_L\theta_L + f_H\theta_H)]\,\hat{Q}^{\frac{1}{2}} - \frac{1}{2}\beta(p^L\delta^L + p^H\delta^H)\hat{Q}^2 \,.$$

The lower β, the lower are the costs of emission reduction under the quantity mechanism, so that $S(\theta, \delta, \hat{Q}, \beta)$ is strictly decreasing in β.

For the parameters in Figure 2.1, if β is close to the ex ante expected value, the price instrument leads to higher expected surplus than the quantity mechanism. But if β differs strongly from the ex ante expected β^*, the quantity mechanism can lead to higher expected total surplus.

2.8 Concluding Remarks

The analysis has combined the classical price and quantity instruments to regulate externalities with a mechanism design framework. I have shown that the presence of private information made it impossible to reach efficient outcomes, when price and quantity instruments were used to regulate externalities. The analysis has established a link between unconstrained mechanism design that reached efficient outcomes and price and quantity instruments. I have shown that although price and quantity instruments lack the flexibility of the unconstrained mechanism, they are durable. The unconstrained mechanism was unable to Pareto-improve upon the classical instruments.

This chapter has abstracted from business-generating activities of firms. Instead, I have isolated the interaction between externality regulation and the design of optimal mechanisms. As a result, the analysis has shown that an optimal unconstrained mechanism cannot improve upon existing price and quantity instruments. Further, I did abstract from consumers responding to the externality. Such problems would likely strengthen the regulation with an unconstrained mechanism, as uncertainty is even higher to misspecify the instruments on the ex ante stage. It is beyond that chapter to provide a game-theoretic analysis of the relationship between consumers and firms when both groups are able to reduce externalities.

Appendix 2.A Proofs

Proof of Proposition 2.1. *Only if – part.* Given a provision rule, Lemmata 1.7-1.12 in the Appendix 1.A imply that the minimal firm revenue that is possible in the presence of firms' interim participation constraints equals

$$\mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\theta,\delta))\right] .$$

If this expression is bigger than $\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(Q(\theta,\delta))\right]$, budget balance cannot be achieved. Thus, constraint efficiency cannot be achieved.
If – part.
Step 1. I need to show that if

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(Q(\theta,\delta))\right] \geq \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\theta,\delta))\right] ,$$

then, given the provision rule $(q_j)_{j=1}^{m}$, $(r_j)_{j=1}^{m}$ can be chosen such that for all j, the incentive compatibility constraints and interim participation constraints are satisfied.

By the arguments in the proof of Lemma 1.11 (see Appendix 1.A, Chapter 1), the participation constraints are satisfied if $R_j(\delta^r) \geq \underline{\pi}_j$. Since monotonicity constraints $C_j(\delta^k) \leq C(\delta^{k-1})$ are satisfied for all j and all l, Lemmata 1.8 and 1.9 in Appendix 1.A imply that incentive compatibility holds if expected transfers are chosen such that all local upward incentive compatibility constraints are binding and that the expected transfer of firm j are equal to

$$\mathbb{E}[t_j(\theta,\delta)] = \mathbb{E}[(\delta_j + h(\delta_j))c(q_j(\theta,\delta))] + \underline{\pi}_j .$$

Step 2. It needs to be shown that if

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(Q(\theta,\delta))\right] \geq \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\theta,\delta))\right] ,$$

then, given the provision rule Q, $(t_1)_{i=1}^{n}$ can be chosen such that for all i, the incentive compatibility constraints are satisfied.

Step 2.1 I introduce a pricing schedule in combination with a lump-sum tax (θ), so that for each i and l,

$$T_i(\theta^l) = \theta + s(B_i(\theta^L)) ,$$

where $s : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a non-decreasing schedule with $s(0) = 0$. I will show that this pricing mechanism satisfies incentive compatibility constraints. The proof follows Bierbrauer (2011a).

Suppose that the price schedule is not incentive compatible. Then, there exists i, l and k so that

$$\theta^l B_i(\theta^l) - T_i(\theta^l) < \theta^l B_i(\theta^k) - T_i(\theta^k) \ .$$

For each $x \in \{B_i(\theta^0), ..., B_i(\theta^m)\}$, $T_i(x) = \tau + s(x)$, implies that

$$\theta^l B_i(\theta^l) - \tau - s(B_i(\theta^l)) < \theta^l B_i(\theta^k) - \tau - s(B_i(\theta^k)) \ .$$

This contradicts the assumption that $B_i(\theta^l) \in argmax_{x\in\mathbb{R}_+}\{\theta^l x - \tau - s(x)\}$. Hence, the assumption that a pricing schedule is not incentive compatible leads to a contradiction and therefore, must be wrong.

Proof of Proposition 2.2.

Cost minimization without participation constraints. I first study the following optimization problem: Given a provision rule $(q_j)_{j=1}^m$ that satisfies the monotonicity constraint $C_j(\delta^{l-1}) \geq C_j(\delta^l)$, for all j and all l, and given expected payments $(R_j(\delta^r))_{j=1}^m$ for firms with a δ^r-costs, I seek to minimize expected costs $\mathbb{E}\left[\sum_{j=1}^m \delta_j c(q_j(\theta,\delta))\right]$ subject to the firms' incentive compatibility constraints

$$R_j(\delta^l) - \delta^l C_j(\delta^l) \geq R_j(\delta^k) - \delta^l C_j(\delta^k) \ ,$$

for all j, l and k. For brevity I refer to this problem as $AUX1$. The following proposition characterizes the solution to this problem.

Proposition 2.7. *A solution $(t_j)_{j=1}^m$ to problem AUX1 has the following properties:*

i) For all j and l, the local upward incentive compatibility constraint,

$$R_j(\delta^{l-1}) - \delta^{l-1}C_j(\delta^{l-1}) \geq R_j(\delta^l) - \delta^{l-1}C_j(\delta^l) \ ,$$

is binding, and all other local upward incentive compatibility constraints are not binding.

ii) The expected payoffs equals

$$\mathbb{E}\left[\sum_{j=1}^m (\delta_j + h(\delta_j))c_j(q_j(\theta,\delta))\right] + \sum_{j=1}^m R_j(\delta^r) \ .$$

iii) The expected firm surplus equals

$$\mathbb{E}\left[\sum_{j=1}^m h(\delta_j)c_j(q_j(\theta,\delta))\right] + \sum_{j=1}^m R_j(\delta^r) \ .$$

The proof follows from Lemmata 1.7-1.12, see Appendix 1.A.

Cost minimization with participation constraints. I define problem AUX2 as follows: Given a provision rule $(q_j)_{j=1}^m$ that satisfies the monotonicity constraint $C_j(\delta^{l-1}) \geq C_j(\delta^l)$, for all j and all l, I seek to minimize expected costs $\mathbb{E}\left[\sum_{j=1}^m \delta_j c(q_j(\theta,\delta))\right]$ subject to the firms' incentive compatibility constraints (IC_F) and the participation constraint

$$R_j(\delta^r) - \delta^r C_j(\delta^r) \geq \underline{\pi}_j \ ,$$

for all j, and l.

Proposition 2.8. *A solution $(r_j)_{j=1}^m$ to problem AUX2 has the following properties:*

i) It has the properties stated in Proposition 2.7.

ii) The participation constraint

$$R_j(\delta^r) - \delta^r C_j(\delta^r) \geq \underline{\pi}_j \ ,$$

is binding for every firm j, whereas all other participation constraints are not binding.

Note that if I modify problem *AUX1* so that the payments $(R_j(\delta^r))_{j=1}^m$ can be freely chosen subject to the participation constraint for δ^r-types, $R_j(\delta^r) \leq \underline{\pi}$, for all j, then the solution will be such that $R_j(\delta^r)) = \underline{\pi}_j$, for all j. To complete the proof it therefore suffices to use the results of Lemmata 1.7-1.12, see Chapter 1.

From Lemma 1.2 in Chapter 1, I know that consumers' incentive compatibility constraints in (IC_C) are satisfied if for all i, $B_i(\theta^l) \geq B_i(\theta^{l-1})$, for all $l > 1$.

I consider the problem of maximizing expected surplus $S(\theta, \delta)$ subject to firms' and consumers' incentive compatibility constraints, firms' participation constraints and the resource requirement in (BC).

I will show that the solution to this problem fulfills property iii) stated in Proposition 2.2.

Step 1. At the solution to the problem, the budget constraint (BC) has to be biding. Otherwise, expected payments of consumers could be decreased and thereby increasing expected surplus $S(\theta, \delta)$. Expected surplus can be written as

$$\mathbb{E}\left[\sum_{i=1}^n \theta_i b\left(\sum_{j=1}^m q_j(\theta,\delta)\right) - \sum_{j=1}^m \delta_j c(q(\theta,\delta))\right] \ ,$$

and the optimal provision rule $(q_j(\theta,\delta))_{j=1}^m$ maximizes this expression. The maximization yields the Samuelson rule, property iii) in Proposition 2.2, when many firms reduce parts of the externality

$$\sum_{i=1}^{n} \theta_i b' \left(\sum_{j=1}^{m} q_j^*(\theta, \delta) \right) = \delta_j c'(q_j^*(\theta, \delta)) \; .$$

Step 2. It is easily verified that externality reduction to the Samuelson rule implies that for all i and all l, the monotonicity constraints $B_i(\theta^l) \geq B_i(\theta^{l-1})$ are satisfied. This implies that consumers' incentive compatibility constraints are satisfied. Further, the Samuelson rule implies that for all j, and all k, the monotonicity constraint $C_j(\delta^{k-1}) \geq C_j(\delta^k)$ is satisfied. Therefore, (PC_F) and (IC_F) are satisfied, see Proposition 2.7 and 2.8.
i) Consumer surplus is

$$U(\theta, \delta) = \mathbb{E}\left[\sum_{i=1}^{n} (\theta_i b(Q^*(\theta, \delta))) - \sum_{i=1}^{n} t_i(\theta, \delta)) \right] \; .$$

Consumers' transfers need to cover firms' expected payoffs, derived in Proposition 2.7 ii), so that

$$U(\theta, \delta) = \mathbb{E}\left[\sum_{i=1}^{n} (\theta_i b(Q^*(\theta, \delta))) - \sum_{j=1}^{m} (\delta_j + h(\delta_j)) c(q_j^*(\theta, \delta)) \right] - \sum_{j=1}^{m} \underline{\pi}_j \; .$$

ii) Expected firm profits are derived in Proposition 2.7 iii).

Proof of Proposition 2.3. *Step 1.* Externality reduction according to the first order condition

$$s - \delta_j c'(q_j^p(\delta_j, s)) \leq 0$$

implies that for all j, and all k, the monotonicity constraint $C_j(\delta^k) \leq C_j(\delta^{k-1})$ is satisfied. I know that (IC_F) and (PC_F) are satisfied, see Appendix 1.A.

Step 2. The mechanism designer has expectation about the amount every firm produces, given he announced s. The expected total amount of externality reduction is determined as an increasing function of the sum of contributions:

$$\mathbb{E}[Q(\delta, s)] = \mathbb{E}\left[\sum_{j=1}^{m} q_j(\delta_j, s) \right] \; .$$

The expected resource costs of externality reduction are $R(\delta, s) = s\mathbb{E}[Q^p(\delta, s)] + \sum_{j=1}^{m} \underline{\pi}_j$. Making use of (BC), expected surplus can be written as

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b(Q(\delta, s)) \right] - R(\delta, s) \; ,$$

which establishes part i) of Proposition 2.3.

This implies as well that

$$R(\delta, s) - \mathbb{E}\left[\sum_{j=1}^{m} \delta_j c(q_j(\delta_j, s))\right] ,$$

which is the expression for firm profits in part ii).

Expected budget balance implies that expected surplus can be written as

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b(Q(\delta, s)) - \sum_{j=1}^{m} \delta_j c(q_j(\delta_j, s))\right]$$

and the optimal price s maximizes this expression. The solution s^* to this problem is such that the following first order condition is satisfied:

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'\left(Q(\delta, s)\right) \frac{\partial Q(\delta, s)}{\partial s} - \sum_{j=1}^{m} \delta_j c'\left(q_j(\delta_j, s)\right) \frac{\partial q_j(\delta_j, s)}{\partial s}\right] = 0$$

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'\left(Q(\delta, s)\right) \frac{\partial Q(\delta, s)}{\partial s} - \sum_{j=1}^{m} s \frac{\partial q_j(\delta_j, s)}{\partial s}\right] = 0$$

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'\left(Q(\delta, s)\right) \frac{\partial Q(\delta, s)}{\partial s} - s \frac{\partial Q(\delta, s)}{\partial s}\right] = 0$$

$$\mathbb{E}\left[\sum_{i=1}^{n} \theta_i b'\left(Q(\delta, s)\right)\right] = s$$

Proof of Proposition 2.4. The proof follows the proof in Proposition 2.2 i) and ii).

Proof of Proposition 2.5. I consider the problem of maximizing expected surplus $S(\theta, \delta, \hat{Q})$ subject to (IC_F) (PC_F), (IC_C), (BC) and additionally taking the constraint

$$\hat{Q} = \sum_{j=1}^{m} q_j(\delta, \hat{Q}) \ .$$

into account, i.e., the quantity goal needs to be met.

Step 1. At the solution to the auxiliary problem (BC) has to be biding. Otherwise, expected payments of consumers could be decreased and with that surplus increased.

Making use of (BC), expected surplus can be written as

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b\left(\sum_{j=1}^{m} q_j(\delta,\hat{Q})\right)-\sum_{j=1}^{m}\delta_j c(q_j(\delta,\hat{Q}))\right] \; , \qquad (B)$$

and the optimal provision rule $\hat{Q} = \sum_{j=1}^{m} q_j(\delta,\hat{Q})$ maximizes this expression. This implies that the objective in (B) can be rewritten as:

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(\hat{Q})-\sum_{j=1}^{m}\delta_j c(q_j(\delta,\hat{Q}))\right] \; .$$

The maximization of expression (B) yields to a Samuelson rule that holds in expectation, property iii) in Proposition 2.5, when many firms reduce parts of the externality

$$\sum_{i=1}^{n}\theta_i b'(\hat{Q}) = \mathbb{E}\left[\sum_{j=1}^{m}\delta_j \frac{\partial c(q(\delta,\hat{Q}))}{\partial q_j(\delta,\hat{Q})}\frac{q_j(\delta,\hat{Q})}{\hat{Q}}\right] \; .$$

Step 2. It is easily verified that externality reduction to the Samuelson rule implies that for all i and all l, the monotonicity constraints $B_i(\theta^l) \geq B_i(\theta^{l-1})$ are satisfied. This implies that consumers' incentive compatibility constraints are satisfied. Further, the Samuelson rule implies that for all j, and all k, the monotonicity constraint $C_j(\delta^{k-1}) \geq C_j(\delta^k)$ is satisfied. Therefore, (PC_F) and (IC_F) are satisfied.

Proof of Proposition 2.6.

i) It follows from Propositions 2.2, 2.3 and 2.5 that the ex post efficient amount of externality reduction under the unconstrained mechanism does neither coincide with the introduced quantity nor the amount of externality reduction under the price mechanism. Further, under all mechanism marginal costs of emission reduction are equal, so that emission reduction occurs at lowest costs. Hence, if more emission reduction is optimal, then the costs of reduction need to increase as well.

Case 1 Consider the first the case where $Q(\theta,\delta)^* > \hat{Q}$. As more externality reduction is optimal under the unconstrained mechanism, firms' need to receive higher revenues in order to make at least as high profits as under the quantity mechanism. Otherwise firms will not agree to switch from a quantity mechanism to the unconstrained mechanism. Therefore, it needs to hold that $R(Q(\theta,\delta)^*) > R(\hat{Q})$. The lump-sum component τ under the unconstrained mechanism is therefore higher under the unconstrained mechanism than under the quantity mechanism. But then, the consumer with the preference parameter θ^0 will veto against the unconstrained mechanism, as $u(\theta^0, Q(\theta,\delta)^* < u(\theta^0,\hat{Q})$. Thus, the unconstrained mechanism cannot replace a quantity mechanism.

Case 2 Consider the next the case where $Q(\theta,\delta)^* < \hat{Q}^*$. The costs of reducing the optimal amount of externalities are therefore lower than the costs of reducing the amount that is fixed by the quantity mechanism. The revenue requirement under the unconstrained mechanism is lower than the revenue requirement under the quantity mechanism ($R(\hat{Q}^*) > R(Q(\theta,\delta)^*)$. But firm j's profit is increasing in the amount of emission reduction $\pi_j(\delta_j^l, q_j) > \pi_j(\delta_j^l, q_j')$ for $q_j > q_j'$, so that firms prefer the quantity mechanism to the unconstrained mechanism and would therefore veto against the unconstrained mechanism. The arguments for the price mechanism are similar.

ii) Quantity: The optimal emission reduction is characterized in Proposition 2.2. Interim, a firm j expects to reduce

$$\mathbb{E}_{(\theta,\delta_{-j})}\left[q_j(\theta,\delta_{-j},\delta_j)|\delta_j=\delta^l\right] .$$

If $l=r$, the optimal amount of expected emission reduction is lower under the optimal unconstrained mechanism than under the quantity mechanism that is characterized in Proposition 2.5

$$\mathbb{E}_{(\theta,\delta_{-j})}\left[q_j(\theta,\delta_{-j},\delta_j)|\delta_j=\delta^l\right] < \mathbb{E}_{(\theta,\delta_{-j})}\left[q_j(\delta_{-j},\delta_j,\hat{Q})|\delta_j=\delta^l\right] .$$

Under the optimal unconstrained mechanism, as well as the quantity mechanism, firm j's expected profit $\Pi_j(\delta^l)$ is increasing in $Q_j(\delta^l)$. A firm with high costs of emission reduction, $l=r$, vetoes the optimal unconstrained mechanism if

$$\mathbb{E}_{(\theta,\delta_{-j})}\left[r_j(\theta,\delta_{-j},\delta_j)|\delta_j=\delta^l\right] < \mathbb{E}_{(\theta,\delta_{-j})}\left[r_j(\delta_{-j},\delta_j,\hat{Q})|\delta_j=\delta^l\right] .$$

Increasing firms' expected revenues is, however, not possible because a consumer with low preferences, $l=0$, vetoes the optimal unconstrained mechanism if transfers under the optimal unconstrained mechanism are higher than under the quantity mechanism

$$\mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\theta_{-i},\theta_i,\delta)|\theta_i=\theta^l\right] < \mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\delta,\hat{Q})|\theta_i=\theta^l\right] .$$

Price: In order to make the optimal unconstrained mechanism Pareto improve upon the price mechanism, the expected profits under the unconstrained mechanism need to be higher than the profits under the price mechanism

$$\mathbb{E}_{(\theta,\delta_{-j})}\left[\pi_j(\theta,\delta_{-j},\delta_j)|\delta_j=\delta^l\right] \geq s^* q_j(\delta_j,s^*) - \delta_j c(q_j(\delta_j,s^*)) .$$

The unconstrained mechanism as well as the price mechanism achieve production efficiency. Thus, expected firm profits under the unconstrained mechanism can only be higher than under the price mechanism if

$$\mathbb{E}_{(\theta,\delta_{-j})}\left[\sum_{j=1}^{m}\left[q_j(\theta,\delta_{-j},\delta_j)|\delta_j=\delta^l\right]\right] > Q(\delta,s^*) .$$

Increasing firms' expected revenues is, however, not possible because a consumer with low preferences, $l = 0$, vetoes the unconstrained unconstrained mechanism if transfers under the unconstrained mechanism are higher than under the price mechanism

$$\mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\theta_{-i},\theta_i,\delta)|\theta_i=\theta^l\right] < \mathbb{E}_{(\theta-i,\delta)}\left[t_i(\delta,s^*)|\theta_i=\theta^l\right] \ .$$

3

Externality regulation and distributional concerns

3.1 Introduction

This chapter contributes to the theory of externality regulation. It looks at a model in which firms can reduce environmental damages, which is beneficial to consumers. The chapter assumes that firms privately observe their reduction costs and consumers privately observe their benefits of environmental damage reduction.

A key question in the literature is to identify the conditions under which the regulation via corrective taxation is preferable to quantity regulation, such as tradable permits. In a world where all costs and benefits are known with certainty, price and quantity mechanisms are equally efficient to regulate the externality. Yet, for the question to be meaningful, this literature focuses on situations in which the policy is chosen under uncertainty about the reduction costs of firms and the benefits of reduction to consumers. Thus, there is uncertainty about the demand (as shaped by the distribution of preferences) and the supply (as shaped by the distribution of reduction costs) for externality reduction. The chosen policy instrument will affect the total surplus, and the equivalence argument does no longer apply.

Most of this literature, however, assumes that there is symmetric information; i.e., there is no explicit modeling of private information and no agent has better information than others. By contrast, this chapter studies externality regulation under asymmetric information. It is based on the assumption that such informational asymmetries are an impediment for the design of an optimal response to externalities. And information asymmetries are present in the context of environmental externalities: A firm's manager or engineer can assess the avoidance costs that emissions will cause for the firm more precisely than the regulatory agency. Further, consumers that are directly affected by emissions can estimate, for example, their heath costs more accurately. This raises the concern that an analysis which makes the assumption that information is symmetric will suffer from an inaccurate assessment of avoidance costs and benefits.

This chapter formalizes a trade-off between the efficiency of regulating externalities, on the one hand, and the distribution of consumer surplus and firm surplus, on the other hand. As a main result, it is shown that if the regulator maximizes the surplus of consumers, who are harmed by the externality, less externality reduction takes place compared to the case when the regulator maximizes total surplus. Further, this paper asks under what circumstances a price mechanism leads to higher surplus than a quantity mechanism and whether this comparative advantage measure is affected by the mechanism designer's surplus measure.

The formal analysis is based on the independent private values model of mechanism design that I introduced in Chapter 1. It has the following features: There are many firms that can reduce emissions; they differ with respect to their technology. On the other side, there is a number of consumers that benefit from the emission reduction, but they differ in their preferences for externality regulation. Furthermore, consumers' transfers need to cover firms' revenues. I assume that the mechanism designer needs to use either a price mechanism, or a quantity mechanism to regulate emissions. This stays in contrast to the literature on the revelation of private information that avoids any a priori assumption on the set of admissible policies. The mechanism designer needs to introduce price and quantity mechanisms before private information is revealed. When using a price mechanism, the mechanism designer introduces a subsidy that firms receive per unit of emission reduction. The total amount of emission reduction is, however, still responsive to changes in technologies because firms decide on their amount of emission reduction after they have learned about their technology and the subsidy. On the other hand, if a quantity mechanism is chosen for externality reduction, the total amount of reduction is certain but uncertainty remains about the cost of reduction.

I impose incentive compatibility constraints and participation constraints for firms. The first constraints assure that firms do not have an incentive to misreport their technologies. The latter constraints assure that no firm makes losses when reducing emissions. I impose no participation constraints for consumers. I assume instead that the regulator can make use of its coercive power and force consumers to pay transfers, even if some consumers do benefit only little from emission reduction. The rationale for this assumption is twofold: First, it has been shown by the literature that emissions will not be reduced if consumers' incentive compatibility constraints and participation constraints need to be respected simultaneously and the number of consumers is large (see e.g. Chapter 1 and Mailath and Postlewaite, 1990). If no externality reduction takes place, the question of whether to use prices or quantities becomes meaningless. And second, it looks empirically plausible that the regulating agency rather forces consumers to pay transfers for externality reduction than letting firms go out of business.

As a benchmark case, I derive the optimal *unconstrained* mechanism to regulate externalities, i.e., a mechanism that does restrict the set of admissible policies, when the mechanism designer's objective is to maximize consumer surplus. I show that less externality reduction takes place, compared to the case where the mechanism designer maximizes total surplus that consists of consumer surplus and firm profits. Put differently, less emissions are reduced when the mechanism designer puts more weight on the surplus of the harmed agents. When total surplus is maximized, the optimal amount of externality regulation has to satisfy an efficiency condition, which is known as the Samuelson rule (Samuelson, 1954). Although consumers and firms have private information, the unconstrained mechanism that maximizes total surplus reaches first-best efficiency. When the mechanism designer puts all weight consumers in the surplus measure, the Samuelson rule no longer needs to be satisfied. Instead, an efficiency condition needs to hold where not only firms' costs of emission reduction enter the optimality condition but also rents they need to receive in order to communicate their private information truthfully. These rents are part of the mechanism designer's surplus measure when he is interested in maximizing total surplus. By contrast, these rents reduce surplus when the mechanism designer maximizes consumer surplus because they increase consumers' transfers. Thus, the optimal reduction level under consumer surplus maximization is distorted downwards, and the unconstrained consumer surplus maximizing mechanism only achieves second-best results.

Next, I introduce a price mechanism. The mechanism designer announces a subsidy that firms receive per amount of emission reduction. When choosing the subsidy, the regulator remains ignorant about the state of demand and supply for externality reduction. The mechanism designer, however, has probabilistic beliefs about demand and supply and hence can assess the performance of a price mechanism. If he acts in the interest of consumers, the emission reduction should occur at lowest costs. The mechanism designer demands externality reduction from many firms. A market structure arises that resembles a monopsony. The subsidy that is paid when consumer surplus is maximized is lower than the subsidy that is paid when total surplus is maximized. As under the optimal mechanism, less externality reduction takes place under the first surplus measure. But as the subsidy is fixed, only the amount of externality reduction can adjust to changes in the economy. In comparison to the unconstrained mechanism that maximizes consumer surplus, a surplus loss occurs when the mechanism designer uses a price mechanism. Thus, a price mechanism that maximizes consumer surplus only leads to third-best efficiency, compared to the unconstrained consumer surplus maximizing mechanism.

Thereafter, I introduce a quantity mechanism that maximizes consumer surplus. The mechanism designer sets the total amount of emission reduction before he knows consumers' preferences and firms' costs. As he acts in the interest of consumers, the quantity mechanism should achieve emission reduction at lowest costs. As firms have private information about their costs, incentive problems can occur: A firm with a better technology tends to overstate its costs to reduce emissions because it profits more from an increase in the revenue requirement. Because incentive constraints and firms' voluntary participation needs to be assured, the optimality condition for emission reduction accounts for firms' rents. In comparison to the optimal consumer surplus maximizing mechanism, the ex ante introduced quantity restriction appears in the optimality condition. That is, the quantity regulation is less flexible to react to changes in costs and preferences. As the price mechanism above, the quantity mechanism that maximizes consumer surplus only achieves third-best efficiency.

While price and quantity mechanisms are often used in real-world applications, they lead to a surplus loss in comparison to an unconstrained mechanism. This raises the question how severe this surplus loss is, and which parameters influence it. Further, if only price and quantity mechanisms are considered, the question, which mechanism leads to higher surplus, becomes important. I show that the mechanism designer's surplus measure affects the comparative advantage of the price mechanism over the quantity mechanism. I present a numerical example to highlight that, in particular, the distribution of firms' cost parameters is crucial to evaluate not only the efficiency but also the distribution of consumer and firm surplus.

The remainder of this chapter is organized as follows: The next section gives a more detailed literature review. Section 3.3 specifies the economic environment. Section 3.4 describes, as a benchmark case, the optimal mechanism that maximizes consumer surplus under the assumption that firms and consumers have private information. Further, I describe the price and the quantity mechanism that maximize consumer surplus. Section 3.5 compares price and quantity mechanisms under efficiency and distributional considerations. The last section contains concluding remarks.

3.2 Related Literature

My research lies at the intersection of two different strands of the literature: On the one hand, the analysis of externality regulation under informational constraints; and on the other hand, the traditional regulation of externalities by means of price and quantity instruments.

Mechanism Design. This chapter builds on the independent private values model of mechanism design theory that was introduced in Chapter 1. I apply this framework to study externalities.

Traditionally, the literature on externality regulation is related to the literature on public good provision and assumes that consumers have private information about their preferences for public good provision. In order to reach surplus-maximizing outcomes, the mechanism designer needs access to consumers' private information. Arrow (1979) and D'Aspremont and Gerard-Varet (1979) show that it is possible to get consumers to reveal their preferences if the mechanism designer's goal is to reach efficient outcomes. If additionally consumers' voluntary participation needs to be assured, then public good provision goes to zero, as the number of consumers get large, see e.g. Rob (1989) and Mailath and Postlewaite (1990). For the context of this chapter, this would mean that externalities are not regulated if a consumer cannot be forced to contribute to the financing of externality reduction, in case that this makes him or her worse off.

I allow not only consumers but also the production sector to have private information and thereby build on results of Baron and Myerson (1982), Bierbrauer (2011a) and Chapter 1. The first paper shows how to regulate a monopolist with unknown costs. The second paper looks at a single firm providing an excludable public good. In contrast to this chapter, these papers do not examine how high the share of public good provision should be for a single firm when multiple firms can provide the public good.

To my knowledge, only one previous paper studies the effects of asymmetric information on the regulation of externalities. Lewis (1996) reviews the literature on incentive regulation and suggests strategies for dealing with asymmetric information in the context of environmental regulation. He looks at a group of firms that differ in their profitability but are equal in the way they pollute the environment. He concludes that firms' private information leads to a pollution level that is different from the efficient one and that taxes should be implemented that vary across firms.

Prices versus Quantities. Several papers in environmental, political and financial economics have described externality regulation and whether to employ price or quantity instruments. My analysis is guided by this literature. The question of prices versus quantities goes back to Weitzman (1974). He showed that both instruments differ in efficiency if information is incomplete but symmetric, i.e., when there is uncertainty about the true demand and supply functions and all agents have the same information. As Weitzman (1974), I analyze price and quantity instruments but assume that uncertainty stems from private information. With that I provide an extension of Weitzman's results to an environment with asymmetric information.

The question of prices versus quantities has especially been applied in the context of environmental economics, see Cropper and Oates (1992) for a survey[1] and in political

[1]Many authors have expanded Weitzman's framework. Yohe (1978) adds output uncertainty, Hoel and Karp (2001) consider multiplicative uncertainties. Stavins (1996) shows that a positive correlation between benefit shocks and environmental costs, favor quantity instruments in a setting with environmental externalities. Goulder et al. (1999) and Pizer (2002) look if a combination of both instruments leads to efficiency improvements in comparison to either prices or quantities alone. Montero (2002)

economy models.[2] Recent papers analyze price and quantity instruments to deal with externalities that arose with the financial crises.[3] As Weitzman (1974), most of these papers assume that uncertainty is due to incomplete information. However, information advantages of a single financial institute with respect to the regulating agency have so far not been examined. The setup of this chapter can be a starting point to investigate incentive effects of bankers, who are better informed about the risk than the government.

Overall, the literature on price and quantity regulation uses total surplus as the efficiency criterion. In this chapter, I additionally consider consumer surplus as efficiency measure and thereby follow the literature on regulation, see Baron and Myerson (1982) and Laffont and Tirole (1993). I will show that the two surplus criteria will lead to different levels of emission reduction and affect the comparison of prices versus quantities.

3.3 The economic environment

3.3.1 Firms

There is a set of firms, $J = \{1, ..., m\}$ that can reduce emissions. The emission reduction q_j leads to costs $\delta_j c(q_j)$, where δ_j is a cost characteristic of the firm that is unknown to the mechanism designer. δ_j belongs to the finite ordered set $\Delta = \{\delta^1, ..., \delta^r\}$ of possible cost parameters. I assume that $\delta^l < \delta^{l+1}, \forall\, l \in \{1, ..., r-1\}$, so that firms with a lower index face lower costs. The profit of firm j is given by

$$\pi_j(\delta_j, t_j, q_j) = r_j - \delta_j c(q_j) \ ,$$

where $c(\cdot)$ is an increasing and convex cost function, satisfying $c(0) = 0 \lim_{q\to 0} c'(q) = 0$ and $\lim_{q\to\infty} c'(q) = \infty$, where the Inada conditions avoid corner solutions. r_j is firm j's revenue.

Each firm privately observes its type δ_j. For all other agents δ_j is a random variable with support Δ and probability distribution $(p^1, ..., p^r)$. I denote by p^l that $\delta_j = \delta^l$

compares both instruments when enforcement is incomplete and Kelly (2005) studies price and quantity regulations in a general equilibrium model.

[2]Boyer and Laffont (1999) recast the question of instrument choice in a formal political economy model of environmental economics and examine the influence of voting rules on the environmental policy. There exists as well a strand of the policy literature that studies the instrument choice in environmental settings. If rent seeking might occur, then taxes are rarely applied in environmental policies (see e.g. Dijkstra, 1998). Damania (1999) shows that emission taxes are more likely when parties act in the interest of the environment, whereas quantities are likelier when interest groups are present. And Alesina and Passarelli (2014) show that a majority might prefer a different instrument than the social planner.

[3]Keen (2011) considers failure externalities and bail-out externalities and evaluates price and quantity instruments. Perotti and Suarez (2011) discuss both instruments when liquidity regulation generates negative systemic risk externalities. Stein (2012) examines monetary policy instruments to regulate negative externalities in the financial sector in order to help assuring financial stability. Acharya et al. (2010) study Pigouvian taxation to internalize systemic risk-taking behavior of banks.

and by $P(\delta^l)$ the probability that $\delta_j > \delta^l$. This distribution is assumed to be common knowledge. The random variables $(\delta_j)_{j\in J}$ are assumed to be independently and identically distributed (*i.i.d.*). The vector of all firms' technology parameters is denoted by $\delta = (\delta_1, ..., \delta_m)$ and δ_{-j} lists all technology parameters except δ_j. I impose a monotone hazard rate assumption: The function

$$\delta_j \mapsto h(\delta_j) := \frac{P(\delta_j)}{p(\delta_j)}$$

is assumed to be non-decreasing.

3.3.2 Consumers

There is a finite set of consumers, $I = \{1, ..., n\}$. The preferences of consumer i are given by the utility function

$$u_i(\theta_i, Q, t_i) = \theta_i b(Q) - t_i \ ,$$

where $Q = \sum_{j=1}^{m} q_j \in \mathbb{R}_+$ denotes the overall emission reduction. The utility depends on a taste parameter θ_i that describes consumer i's marginal valuation for emission reduction. For each i, θ_i belongs to a finite ordered set $\Theta = \{\theta^0, \theta^1, ..., \theta^s\}$, with $\theta^0 = 0$. I assume that $\theta^l < \theta^{l+1} \ \forall \ l \in \{0, ..., s-1\}$, etc. The monetary transfer of consumer i is denoted by t_i.

Each consumer i privately observes his preference parameter θ_i. From the perspective of all other agents it is a random variable with support Θ and probability distribution $(f^0, ..., f^s)$. I denote by f^l that $\theta_i = \theta^L$ and by $F(\theta_i^l)$ the probability that $\theta_i > \theta^l$. The random variables $(\theta_i)_{i\in I}$ are *i.i.d.* The vector of all consumers valuation parameters is denoted by $\theta = (\theta_1, ..., \theta_n)$. θ_{-i} lists all taste parameters except θ_i. I impose a monotone hazard rate assumption, so that the function

$$\theta_i \mapsto g(\theta_i) := \frac{1 - F(\theta_i)}{f(\theta_i)}$$

is assumed to be non-increasing.

3.3.3 Mechanism

I use a mechanism design approach to characterize the amount and the pricing of emission reduction. A social choice function or direct mechanism consists of a transfer and production rule for each firm j and a level of externality reduction and transfer rule for each consumer i. $r_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}$ specifies j's revenue as a function of the vector of agents reports; analogously the function $q_j : \Theta^n \times \Delta^m \mapsto \mathbb{R}_+$ characterizes j's emission reduction. $t_i : \Delta^n \times \Delta^m \mapsto \mathbb{R}$ characterizes consumer i's payment as a function of the vector of consumers' preference parameters and firms' cost parameters. Consumers benefit from the total emission reduction $Q(\theta, \delta) = \sum_{j=1}^{m} q_j(\theta, \delta)$.

By referring to the revelation principle (Myerson, 1985), I focus on direct mechanisms so that the truthful revelation of preferences for externality reduction by consumers, and technologies by firms, is a Bayes-Nash equilibrium.

Incentive compatibility constraints. Truth-telling is a best response for firm j if for all $\delta^l \in \Delta$ and all $\delta^k \in \Delta$

$$R_j(\delta^l) - \delta^l C_j(\delta^l) \geq R_j(\delta^k) - \delta^l C_j(\delta^k), \qquad (IC_F)$$

where $R_j(\delta^k) \equiv \mathbb{E}_{(\theta,\delta_{-j})}\left[r_j(\theta, \delta_{-j}, \delta_j^k)\right]$ is the expected transfer for firm j, in case of reporting δ^k, given that all other firms and consumers reveal their preferences to the mechanism designer. Analogously, $C_j(\delta^k) \equiv \mathbb{E}_{(\theta,\delta_{-j})}\left[c(q_j(\theta, \delta_{-j}, \delta_j^k))\right]$ are the expected production costs of firm j when announcing δ^k. The expectation operator $\mathbb{E}_{(\theta,\delta_{-j})}$ indicates that the realization of δ_j is known when deriving this expectation.

In the same way, truth-telling is a best response for consumer i if for all $\theta^l \in \Theta$ and all $\theta^k \in \Theta$

$$\theta^l B_i(\theta^l) - T_i(\theta^l) \geq \theta^l B_i(\theta^k) - T_i(\theta^k), \qquad (IC_C)$$

where $B_i(\theta^k) \equiv \mathbb{E}_{(\theta_{-i},\delta)}\left[b(Q(\theta_{-i}, \theta_i^k, \delta))\right]$ is the expected benefit of emission reduction, conditioning on consumer i reporting θ^k, given that all other consumers and firms reveal their preferences to the mechanism designer and $T_i(\theta^k) \equiv \mathbb{E}_{(\theta_{-i},\delta)}\left[t_i(\theta_{-i}, \theta_i^k, \delta)\right]$ are the expected payments, in case of reporting a preference parameter of θ^k.

Participation constraints. A direct mechanism is individually rational if for each firm j and for all $\delta^l \in \Delta$

$$R_j(\delta^l) - \delta^l C_j(\delta^l) \geq \underline{\pi}_j, \qquad (PC_F)$$

where $\underline{\pi}_j$ denotes a lower bound for the profit of firm j. The interpretation of these participation constraints is as follows: One can think of the direct mechanism as a mechanism that replaces a status quo outcome. The mechanism designer requires an unanimous consent of all firms. Hence, $\underline{\pi}_j$ can be understood as firm j's payoff in the status quo.

I assume that the mechanism designer can make use of its coercive power and force consumers to pay a transfer for the externality reduction. Thus, voluntary participation for consumers is not required. I make this assumption because the literature has shown that no externality reduction takes place if consumers' incentive compatibility constraints and participation constraints need to be respected simultaneously. Thus, the question of which mechanism to use for externality regulation would become meaningless without assuming coercive power.

Budget constraint. Furthermore, the mechanism designer's budget constraint needs to hold, where consumers expected transfers need to cover firms' expected revenues

$$\mathbb{E}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] \geq \mathbb{E}\left[\sum_{j=1}^{m} r_j(\theta,\delta)\right] . \qquad (BC)$$

The objective function. I will consider two alternatives in what follows. Expected total surplus that is generated by a social choice function is given by

$$S((r_j)_{j\in J},(q_j)_{j\in J},(t_i)_{i\in I}) = \Pi((r_j)_{j\in J},(q_j)_{j\in J}) + U(Q,(t_i)_{i\in I}) \ ,$$

where firms' expected revenues are

$$\Pi((r_j)_{j\in J},(q_j)_{j\in J}) = \mathbb{E}\left[\sum_{j=1}^{m} (r_j(\theta,\delta) - \delta_j c(q_j(\theta,\delta)))\right] \ ,$$

and expected consumer surplus is

$$U(Q,(t_i)_{i\in I}) = \mathbb{E}\left[\sum_{i=1}^{n} (\theta_i b(Q(\theta,\delta)) - t_i(\theta,\delta))\right] \ .$$

In Chapter 2, I derived results when the regulator is interested in maximizing total surplus. This surplus measure is a target that is mostly used when comparing the efficiency of price and quantity mechanisms. Expected total surplus, however, does not account for distributive consideration. The assumption that a consumer's utility is linear in transfer t_i and a firm's profit is linear in the revenue r_j gives rise to an externality reduction level that does not depend on which agent should be able to realize which surplus level. The criterion of total surplus maximization pins down Q and $(q_j)_{j\in J}$. This criterion, however, takes a particular stand on distributive concerns: Firm profits and consumer surplus receive the same weight in the mechanism designer's objective function, hence distributive concerns are neglected.

A possibility to formalize distributive consideration is to look at an objective function that maximizes the surplus of one group in the economy. A key question of this chapter is to specify the mechanism that maximizes consumer surplus and comparing it to the outcomes of expected total surplus maximization.[4]

[4]I do not consider the case of maximizing firm profit because it would no longer be desirable to restrict firms revenues, and the regulation of externalities would become uninteresting. As consumers' participation constraints are neglected, firms' revenues can be made arbitrarily high, such that a firm's profit maximization problem does not have a solution.

3.4 Externality regulation under asymmetric information

In response to the non-optimality of the market equilibrium in the presence of emissions, various instruments have been proposed to correct for the distortions. In this section I introduce an unconstrained mechanism and two specific mechanisms that are widely used in real-world applications: A quantity mechanism; that is to say a mechanism that achieves the same level of emission reduction, independently of the realized demand and supply of emission reduction, and a price mechanism that achieves the same marginal costs of emission reduction for any realization of supply and demand.

To evaluate the efficiency of price and quantity mechanisms, an unconstrained mechanism to regulate emissions is derived. The advantage of that mechanism is that it specifies the emission level for every realization of firms' technologies and consumers' preferences. This makes the unconstrained mechanism more flexible than price and quantity mechanisms, that fix either the price paid for emission reduction or the total emission level before uncertainty is resolved.

In this section, I compare the outcome that is obtained if total surplus is maximized to the outcome that is obtained if consumer surplus is maximized. As will become apparent, with a regulator that is interested in maximizing consumer surplus less emissions are reduced than when he is interested in maximizing total surplus. This result is independent of the mechanism being used.

3.4.1 Optimal mechanism design as a benchmark case

As a benchmark case, I first introduce the unconstrained mechanism. There is an ex ante stage. At this stage the mechanism designer wants to define the mechanism for emission reduction. The mechanism designer acts in the interest of consumers. He has information about the distribution of preferences and technologies. He does, however, not know the private information of a single firm or a single consumer. The designer employs firms' incentive compatibility and participation constraints, so as to truthfully elicit firms' private information. Additionally, he requires that the budget is balanced in expectation, so that consumers' expected transfers cover firms' expected revenues.

Formally, I require that a social choice function is chosen with the objective to maximize consumer surplus subject to the incentive compatibility constraints in (IC_F), firms' participation constraints (PC_F) and the budget constraint in (BC). I first consider the problem to derive the mechanism that reaches the emission reduction at minimal costs, and additionally respecting firms' constraints. In a second step, I derive the emission level that maximizes consumer surplus. Proposition 3.1 specifies the solution of the first problem.

Proposition 3.1. *Suppose that the monotone hazard rate assumption holds (h is a non-decreasing function). The optimal mechanism that reaches emission reduction at minimal costs and that satisfies (IC_F), (PC_F) and (BC), leads to the following resource requirement*

$$R(\theta, \delta) = \mathbb{E}\left[\sum_{j=1}^{m} (\delta_j + h(\delta_j))\, c(q_j(\theta, \delta))\right] + \sum_{j=1}^{m} \underline{\pi}_j \ .$$

For a formal proof of this argument, see Chapter 1.[5] As firms have private information about how costly it is to reduce emissions, and as the expected profit of a single firm is not allowed to be negative, firms can achieve information rents. These information rents enter the resource requirement. The cost function of a complete information environment is replaced by firm j's virtual cost function $(\delta_j + h(\delta_j))\, c(q_j(\theta, \delta))$ for emission reduction. Virtual costs can be interpreted as the minimal revenue that type δ_j of firm j needs to receive for the emission reduction in an environment where information is asymmetrically distributed. The costs for emission reduction are increased by the hazard rate $h(\delta_j)$ that accounts for firms' private information.

After having solved the problem of reaching emission reduction at lowest costs for consumers, I specify now the optimal level of emission reduction.

Proposition 3.2. *For any reported preferences θ and technologies δ, the emission reduction level of firm j, that maximizes consumer surplus U, which I denote by $q_j^{**}(\theta, \delta)$, has to satisfy the following first order condition, for all firms j*

$$\sum_{i=1}^{n} \theta_i b'\left(\sum_{j=1}^{m} q_j^{**}(\theta, \delta)\right) = (\delta_j + h(\delta_j))\, c'\left(q_j^{**}(\theta, \delta)\right) \ .$$

Proposition 3.2 shows that in order to maximize consumer surplus, the sum of marginal benefits of emission reduction needs to be equal to the virtual marginal costs. For the determination of the optimal emission reduction $Q^{**}(\theta, \delta) = \sum_{j=1}^{m} q_j^{**}(\theta, \delta)$, firms' information rents play a major role. This is different when the regulator is interested in maximizing total surplus, as the following Remark shows.[6]

Remark 3.1. *For any reported preferences θ and technologies δ, the optimal emission reduction, that maximizes total surplus $S(\theta, \delta)$, which I denote by $q_j^{*}(\theta, \delta)$, has to satisfy the following first order condition, for all firms j*

$$\sum_{i=1}^{n} \theta_i b'\left(\sum_{j=1}^{m} q_j^{*}(\theta, \delta)\right) = \delta_j c'\left(q_j^{*}(\theta, \delta)\right) \ .$$

[5] See Appendix 1.A, for a formal proof.
[6] See Appendix 2.A, for a formal proof.

Under the total surplus maximizing mechanism, a condition needs to be satisfied that is an adaption of the Samuelson rule, see Samuelson (1954), to my setup. It states that the the Pareto-efficient emission reduction is reached when the sum of marginal benefits equals the marginal costs of emission reduction. Although firms' costs are private information, first-best efficiency can be reached. The total surplus maximizing emission level is undistorted.

Proposition 3.2 illustrates that under consumer surplus maximization, however, a Samuelson rule no longer holds and first-best results are out of reach. Firms' information rents are no longer part of the mechanism designer's surplus measure. The consumer surplus maximizing mechanism distorts the externality reduction level downwards, in order to realize surplus gains for consumers from redistribution surplus from firms to them. There is a trade-off between higher marginal benefits if emission reduction is expanded and higher costs that incorporate both marginal costs and information rents of the production sector. Only second-best efficiency can be reached.

3.4.2 Price regulation

One way to regulate externalities, in real-world applications, is to impose a subsidy to correct the non-optimality of the market equilibrium. Under a price regulation, a subsidy to reduce emissions is fixed before uncertainty about preferences and costs is resolved. I seek to determine the optimal subsidy that maximizes consumer surplus. The analysis is divided into two parts. I begin by introducing the profit maximization problem of a single firm. In a second step, I solve for the optimal consumer surplus maximizing subsidy.

Optimal emission reduction. The regulator specifies a subsidy s and every firm j maximizes its payoffs, that is

$$\pi_j(\delta_j, s) = sq_j(\delta_j, s) - \delta_j c(q_j(\delta_j, s)) \ ,$$

where $q_j(\delta_j, s)$ is the amount of emissions firm j reduces if subsidy s is announced and if its technology parameter is δ_j. Firm j, therefore, chooses $q_j(\delta_j, s)$ according to the first order condition $s - \delta_j c'(q_j(\delta_j, s)) \leq 0$, with equality for $q_j(\delta_j, s) > 0$, so that the subsidy is equal to the marginal cost of emission reduction for all j. This shows, in particular, that marginal costs of emission reduction are the same for all firms. This will be different under the quantity mechanism, as we will see in the following subsection.

Note further, that participation and incentive constraints in (PC_F) and (IC_F), are satisfied. If firm j had deviated from reducing $q_j(\delta_j, s)$, its profits would decrease. Further, voluntary participation is fulfilled, as each firm can decide to not provide any emission reduction if this leads to negative profits.

With a price mechanism, the resource constraint takes the following form

$$R(\delta, s) = s\mathbb{E}\left[Q\left(\delta, s\right)\right] + \sum_{j=1}^{m} \underline{\pi}_j \ ,$$

where $\mathbb{E}[Q\left(\delta, s\right)] = \mathbb{E}\left[\sum_{j=1}^{m} q_j(\delta_j, s)\right]$ denotes the total expected emission reduction, when subsidy s is paid.

The production sector's profits from emission reduction, when a price regulation is used, are

$$\Pi(\delta, s) = s\mathbb{E}\left[Q\left(\delta, s\right)\right] - \mathbb{E}\left[\sum_{j=1}^{m} \delta_j c\left(q_j\left(\delta_j, s\right)\right)\right] + \sum_{j=1}^{m} \underline{\pi}_j \ .$$

Derivation of the optimal price. The price mechanism needs to maximize consumer surplus, subject to the revenue constraint. I denote the consumer surplus maximizing subsidy by s^{**}.

Proposition 3.3. *For any θ and δ, the optimal subsidy s^{**}, that maximizes consumer surplus $U(\theta, \delta, s)$, has to satisfy the following first order condition*

$$\mathbb{E}\left[\left\{\sum_{i=1}^{n} \theta_i b'\left(Q\left(\delta, s^{**}\right)\right) - s^{**}\right\} \frac{\partial Q(\delta, s^{**})}{\partial s^{**}}\right] = \mathbb{E}[Q(\delta, s^{**})] \ .$$

The expression in the curly bracket of Proposition 3.3 is known from the total surplus maximizing subsidy.[7] If this expression is zero, the modified Samuelson rule is satisfied. Under consumer surplus maximization the first order condition changes. Here, the difference between the expected sum of marginal benefits and the subsidy has to be multiplied by the marginal change of total emission reduction at s^{**} if the subsidy changes. This expression has to be equal to the total level of emission reduction. The following example introduces some further assumptions, so that a comparison to the total surplus maximizing subsidy is simplified.

Example: *Assume firms with a quadratic cost function, $c(q_j) = \frac{1}{2}q_j^2$, and consumers with a concave benefit function, $b(Q) = \sqrt{Q}$. Then, every firm j decides to reduce $q_j = \frac{s}{\delta_j}$, so that $\mathbb{E}[Q(\delta, s)] = \mathbb{E}\left[\sum_{j=1}^{m} q_j(\delta_j, s)\right] = s\mathbb{E}\left[\sum_{j=1}^{m} \frac{1}{\delta_j}\right]$. Hence, $\frac{\partial \mathbb{E}Q(\delta,s)}{\partial s} = \mathbb{E}\left[\sum_{j=1}^{m} \frac{1}{\delta_j}\right]$. The first order condition in Proposition 3.3 then simplifies to*

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \theta_i\right) b'\left(Q\left(\delta, s^{**}\right)\right)\right] = 2s^{**} \ .$$

[7] See Remark 3.2 below.

It is instructive to look at an equivalent formulation in terms of quantity choices. A first order condition can be derived by thinking about the regulator as a monopsonist that decides on the level of emissions it wants to reduce $\mathbb{E}[Q(\delta, s)] \geq 0$, letting the subsidy at which it can buy this amount be given by the inverse supply function $s = (\mathbb{E}[Q])^{-1}$. Using the inverse supply function, the monopsonist's demand function can be stated as

$$max_{E[Q]}\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(Q)\right] - s\left(\mathbb{E}[Q]\right)\mathbb{E}[Q] \; .$$

The optimal emission level must hence satisfy the first order condition

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\theta_i\right) b'(Q)\right] = s'(\mathbb{E}[Q])\mathbb{E}[Q] + s(\mathbb{E}[Q]) \; .$$

The left-hand side of the equation above is the marginal benefit of a differential increase in emissions at $\mathbb{E}[Q]$, while the right-hand side is equal to the marginal resource costs from a differential increase in $\mathbb{E}[Q]$, which is equal to the derivative of the resource costs $\frac{d[s(\mathbb{E}[Q])\mathbb{E}[Q]]}{dE[Q]}$. Thus, marginal resource costs must equal marginal benefits at the monopsonist's optimal emission level. This implies first, that we must have that $s(\mathbb{E}[Q]) < c'(\mathbb{E}[Q])$, so that the subsidy under the monopsony is lower than the marginal costs. And second, that the optimal emission level is below the total surplus maximizing emission level. The cause of that quantity distortion is the regulator's monopsonistic power and the recognition that a reduction in the bought quantities allows it to decrease the subsidy paid on its remaining reductions. This subsidy decrease has an influence on costs and it is captured by the term $s'(\mathbb{E}[Q])\mathbb{E}[Q] + s(\mathbb{E}[Q])$.

In order to compare this price with the consumer surplus maximizing subsidy in Proposition 3.3,I show how the optimal price mechanism looks like if the regulator wants to maximize total surplus.[8]

Remark 3.2. *For any δ and θ, the optimal subsidy s^* that maximizes total surplus $S(\theta, \delta, s)$ has to satisfy the following first order condition*

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\theta_i\right) b'\left(Q\left(\delta, s^*\right)\right)\right] = s^* \; .$$

This condition is a modified Samuelson rule, where the marginal cost of emission reduction s, need to equal the sum of expected marginal benefits. Note that this condition does not coincide with the optimal emission reduction in Remark 3.1. The modified Samuelson rule, needs to be fulfilled only in expectation. The optimal price mechanism

[8]See Chapter 2, for a formal proof.

does not lead to first-best efficient outcomes whenever the regulator's ex ante expectations about preferences and technologies do not coincide with the ex post realization.

The subsidy that is paid when the regulator maximizes consumer surplus is lower than the subsidy paid when the regulator maximizes total surplus. Therefore, the expected total emission reduction is lower under the first surplus measure. Proposition 3.2 highlights, that the regulator's monopsony power has two effects. First, it redistributes surplus away from firms to consumers. Second, it reduces the aggregate expected total surplus, consisting of expected firm profits and expected consumer surplus, as consumers' expected surplus gain is smaller than firms' expected profit loss; a surplus loss that can be understood as a dead weight loss of monopsony.

That is, under a consumer surplus maximizing price mechanism, expected consumer surplus is even lower than under the second-best optimal mechanism, that was introduced in Proposition 3.2. As the price mechanism fixes the subsidy ex ante, it cannot react to changes in preferences and costs by changing the price. The expected consumer utility under a price mechanism is therefore lower than the expected consumer surplus under the optimal mechanism. Thus, the subsidy that satisfies the condition in Proposition 3.3 only leads to 'third-best efficiency'.

3.4.3 Quantity regulation

Instead of a price mechanism, the regulator can use a quantity mechanism to regulate externalities. I think of the quantity mechanism that the regulator chooses the total amount of emission reduction based on his probabilistic beliefs about demand and supply for externality reduction. After the regulator sets the quantity, every firm decides how much it wants to contribute to the total emission reduction. At that point every firm knows its cost parameter. I am interested in the quantity mechanism that maximizes consumer surplus. As consumers' transfers need to cover the expected resource requirement in expectation, and firm profits are not part of the surplus measure, the regulator wants to reach the emission target at minimal resource costs.

Again, I proceed in two steps. First, I treat the optimal quantity as given and solve for the optimal quantity mechanism reaching that aim. In a second step, I solve for the optimal quantity.

Reaching a given target. A quantity mechanism specifies for each firm j a function $q_j : \delta \times \hat{Q} \mapsto q_j(\delta, \hat{Q})$, which gives firm j's contribution to the overall optimal emission reduction as a function of the vector δ of firms' technology parameters and total emission reduction $\hat{Q}$. In addition, a quantity mechanism specifies, for each firm j, a revenue $r_j : \delta \times \hat{Q} \mapsto r_j(\delta, \hat{Q})$, which compensates firm j for its cost of emission reduction. Additional to the constraints that were employed under the optimal mechanism, a quantity constraint needs to be satisfied, that comes from fixing the emission level before uncertainty is resolved, i.e., for all δ,

$$\sum_{j=1}^{m} q_j(\delta, \hat{Q}) = \hat{Q} \; . \tag{QU}$$

The following Proposition characterizes the solution to this problem.

Proposition 3.4. *Suppose that a monotone hazard rate assumption holds (h is a non-decreasing function). The quantity mechanism, which reaches the emission target at minimal resource costs to the regulator, has the following properties*

i) Virtual marginal costs are equalized: For all δ, and for all firms j and j',

$$(\delta_j + h(\delta_j))\, c'\left(q_j\left(\delta, \hat{Q}\right)\right) = (\delta_{j'} + h(\delta_{j'}))\, c'\left(q_{j'}\left(\delta, \hat{Q}\right)\right) \; .$$

ii) The resource requirement of the regulator equals

$$R(\delta, \hat{Q}) = \mathbb{E}\left[\sum_{j=1}^{m} (\delta_j + h(\delta_j))\, c\left(q_j\left(\delta, \hat{Q}\right)\right)\right] + \sum_{j=1}^{m} \underline{\pi}_j \; .$$

iii) The production sectors' expected payoff are

$$\Pi(\delta, \hat{Q}) = \mathbb{E}\left[\sum_{j=1}^{m} h(\delta_j) c\left(q_j\left(\delta, \hat{Q}\right)\right)\right] + \sum_{j=1}^{m} \underline{\pi}_j \; .$$

The Proposition establishes the properties of an optimal quantity mechanism when the emission reduction needs to be reached at lowest costs for consumers. Consumers' transfers not only need to cover the costs of emission reduction but as well firms' information rents. The resource requirement and firm profits depend not only on the cost vector but also on the ex ante introduced quantity goal. Under the quantity mechanism, the amount of emission reduction is fixed ex ante and cannot adapt if preferences or costs are different than ex ante expected. Therefore, the quantity mechanism is less flexible than the optimal mechanism, introduced in Proposition 3.2, where the total amount of emission reduction depends on the realization of preferences and costs.

The term $(\delta_j + h(\delta_j))c'\left(q_j\left(\delta, \hat{Q}\right)\right)$ is firm j's virtual marginal costs of emission reduction under the quantity mechanism, when the regulator decides to reduce emissions by the amount $\hat{Q}$. Remember that the virtual marginal costs can be interpreted as the minimal amount that firm j of type δ needs to receive as a transfer from the regulator in order to reduce its emissions by the amount $q_j\left(\delta, \hat{Q}\right)$, in the presence of incentive compatibility and participation constraints and the quantity constraint. The fact that virtual costs are equalized among firms guarantees that firms with the same technology need to reduce emissions by the same amount but firms with better technology reduce more emissions than firms with bad technology. The emission reduction is undistorted for the firm with the best technology. But emissions are distorted downwards for firms with bad technology due to the hazard rate. Firms with good technology can make use of their information advantage, so that they need to reduce less emissions compared to a case where the regulator had complete information about technologies. A regulator that is interested in reaching the emission reduction at minimal costs for consumers should therefore demand for equal virtual costs of firms. If instead marginal costs where equated, expected costs would increase and with that consumers' transfers.

The optimal target. I assume that the externality target is formulated before consumers' preferences are drawn. I also assume that any resource requirement, that the regulator faces, has to be met by taxing consumers and abstracts from outside founding options. The regulator has expectations about consumers preferences and defines the optimal amount of emission reduction on basis of its expectation. The true position of the benefit function is, however, uncertain ex ante.

The regulator chooses a mechanism $(q_j, r_j)_{j=1}^m$ to maximize expected consumer surplus subject to firms' incentive compatibility (IC_F) and participation constraints (PC_F), the quantity constraint (QU), and subject to the budget constraint (BC). The following Proposition characterizes the solution to the problem.

Proposition 3.5. *For any θ and any δ, the provision level $\hat{Q}^{**}$ that maximizes U, subject to the constraints in (IC_F), (PC_F), (QU) and (BC) , has to satisfy the following first order condition*

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i\right] b'\left(\hat{Q}^{**}\right) = \mathbb{E}\left[\sum_{j=1}^{m}\left(\delta_j + h(\delta_j)\right) c'\left(q_j\left(\delta, \hat{Q}^{**}\right)\right)\right] ,$$

where $c'(q_j) = \frac{\partial c(q_j(\delta,\hat{Q}^{**})}{\partial q_j(\delta,\hat{Q}^{**})}\frac{\partial q_j(\delta,\hat{Q}^{**})}{\partial \hat{Q}}$.

The first order condition states that the sum of expected preferences times the marginal benefits of emission reduction needs to equal the sum of expected virtual costs when the regulator uses a quantity mechanism to reduce emissions. As the regulator only fixes the quantity, there is still uncertainty about the realization of preferences and costs. For a better comparison, the following Remark shows the first order condition when the regulator is interested in maximizing total surplus.[9]

Remark 3.3. *For any θ and δ, the provision level $\hat{Q}^{*}$ that maximizes $S(\theta, \delta, \hat{Q})$, subject to the constraints in (IC_F), (PC_F), (BC) and (QU), has to satisfy the following first order*

[9]See Chapter 2, for a proof.

condition

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i\right] b'(\hat{Q}^*(\delta,\hat{Q}^*)) = \mathbb{E}\left[\sum_{j=1}^{m}\delta_j c'(q_j(\delta,\hat{Q}^*))\right] ,$$

where $c'(q_j) = \frac{\partial c(q_j(\delta,\hat{Q}^{**})}{\partial q_j(\delta,\hat{Q}^{**})}\frac{\partial q_j(\delta,\hat{Q}^{**})}{\partial \hat{Q}}$.

As under the price mechanism that maximized total surplus, a modified Samuelson rule holds. But now the sum of expected preferences times the marginal benefits of emission reduction needs to be equal to the expected marginal costs. Hence, under a quantity mechanism marginal benefits of emission reduction are certain, but marginal costs and the realization of consumers' preferences remain uncertain.

The amount $\hat{Q}^{**}$, that is implemented if consumer surplus is maximized, is lower than $\hat{Q}^*$, the emission reduction under total surplus maximization because the sum of expected virtual costs in Proposition 3.5 exceeds the sum of expected marginal costs in Remark 3.3. In contrast to the total surplus maximizing emission level, the consumer surplus maximizing emission reduction $\hat{Q}^{**}$ fails to satisfy a Samuelson rule even in expectation. Under consumer surplus maximization firms' information rents are no longer part of the surplus measure and less emission reduction is optimal. As under the price regulation, consumers prefer less emission reduction than the reduction that is optimal from total surplus maximization. Thus, the consumer surplus maximizing quantity mechanism only leads to 'third-best efficiency' as the consumer surplus maximizing price mechanism above.

3.5 Prices versus quantities under asymmetric information

With the results from the previous section, I can now compare the unconstrained mechanism with price and quantity mechanisms, when agents are privately informed about their characteristics. First, I compare all mechanisms ex ante, that is before any uncertainty is resolved and do comparative statics. Second, I compare price and quantity mechanisms ex post, that is when all information is publicly known.

To establish functional forms, it provides helpful to introduce some further assumptions on the economy.

Assumption 3.1. *There is one firm* $J = \{1\}$. *The cost parameter* δ_1 *belongs to the binary set* $\Delta_1 = \{\delta^L, \delta^H\}$. *The cost function is quadratic,* $c(q_1) = \frac{1}{2}q_1^2$.

I denote by p^L the probability of a low cost realization, $p^L := Prob\{\delta_1 = \delta^L\}$ and $p^H = 1 - p^L$ for a high cost realization, accordingly.

Assumption 3.2. *There is one consumer* $I = \{1\}$. *The preference parameter* θ_1 *belongs to the binary set* $\Theta_1 = \{\theta^L, \theta^H\}$. *The benefit function is assumed to be concave,* $b(Q) = \sqrt{Q}$.

I denote by f^L the probability of a low preference realization, $f^L := Prob\{\theta_1 = \theta^L\}$ and $f^H = 1 - f^L$ for a high preference realization, accordingly. These assumptions help to illustrate important aspects of the model. In the following, I skip the index 1, for ease of notation.

Optimal mechanism. The given assumptions yield a model for externality regulation that can be solved analytically. Using Proposition 3.2 and Remark 3.1, the following Corollary derives optimal emission reduction levels when first, consumer surplus is maximized and second, total surplus is maximized, for a given state of the economy, or equivalently, for given values of θ and δ. It also gives explicit formulas for the expected consumer surplus and the expected total surplus.

Corollary 3.1. *Suppose that Assumptions 3.1 and 3.2 hold. The regulator uses an unconstrained mechanism to regulate externalities.*

i) If U is maximized, the optimal emission reduction, for a given state of the economy, is given by

$$Q^{**}(\theta,\delta) = \left(\frac{1}{2}\frac{\theta}{\delta + h(\delta)}\right)^{\frac{2}{3}} .$$

If S is maximized, the optimal emission reduction, for a given state of the economy, is given by

$$Q^{*}(\theta,\delta) = \left(\frac{1}{2}\frac{\theta}{\delta}\right)^{\frac{2}{3}} .$$

ii) If U is maximized, the expected consumer surplus is

$$U(\theta,\delta) = \left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{4}{3}}\right]\left[f^L p^L \left(\frac{(\theta^L)^4}{\delta^L + h(\delta^L)}\right)^{\frac{1}{3}} + f^H p^L \left(\frac{(\theta^H)^4}{\delta^L + h(\delta^L)}\right)^{\frac{1}{3}} + f^L p^H \left(\frac{(\theta^L)^4}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}} + f^H p^H \left(\frac{(\theta^H)^4}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}}\right] - \underline{\pi}_j .$$

iii) If S is maximized, the expected total surplus is

$$S(\theta,\delta) = \left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{7}{3}}\right]\left[f^L p^L \left(\frac{(\theta^L)^4}{\delta^L}\right)^{\frac{1}{3}} + f^H p^L \left(\frac{(\theta^H)^4}{\delta^L}\right)^{\frac{1}{3}} + f^L p^H \left(\frac{(\theta^L)^4}{\delta^H}\right)^{\frac{1}{3}} + f^H p^H \left(\frac{(\theta^H)^4}{\delta^H}\right)^{\frac{1}{3}}\right] .$$

Part i) of Corollary 3.1 shows that for a given state of the economy, emission reduction under total surplus maximization is higher than under consumer surplus maximization, whenever the hazard rate $h(\delta)$ is different from zero. Emissions are distorted downwards under consumer surplus maximization. When the regulator puts all weight on the consumers in the surplus function, firms' information rents are no longer part of the surplus measure. The requirement that emission reduction needs to be obtained with lowest transfers for consumers leads to losses in total surplus. The difference between $S(\theta, \delta)$ and $U(\theta, \delta)$ can hence be interpreted as efficiency loss that occurs due to distributive considerations.

Price mechanism. The following Corollary derives the optimal subsidies for both surplus measures. It also gives explicit formulas for expected consumer surplus and expected total surplus.

Corollary 3.2. *Suppose that Assumptions 3.1 and 3.2 hold. Suppose further that the regulator uses a price mechanism to regulate externalities.*

i) The subsidy under consumer surplus maximization and total surplus maximization are, respectively

$$s^{**} = \frac{\left(\frac{1}{4}(f^L\theta^L + f^H\theta^H)\right)^{\frac{2}{3}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}}}, \qquad \textit{and} \qquad s^* = \frac{\left(\frac{1}{2}(f^L\theta^L + f^H\theta^H)\right)^{\frac{2}{3}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}}}.$$

ii) If U is maximized, then expected consumer surplus is

$$U(\theta, \delta, s^{**}) = \left(f^L\theta^L + f^H\theta^H\right)\left[p^L\left(\frac{s^{**}}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{s^{**}}{\delta^H}\right)^{\frac{1}{2}}\right] - s^{**2}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right) - \underline{\pi}_j .$$

iii) If S is maximized, the expected total surplus is

$$S(\theta, \delta, s^*) = \left(f^L\theta^L + f^H\theta^H\right)\left[p^L\left(\frac{s^*}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{s^*}{\delta^H}\right)^{\frac{1}{2}}\right] - \frac{1}{2}s^{*2}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right) .$$

Similar to the unconstrained mechanism in Corollary 3.1, if consumer surplus is maximized, less emissions are reduced than when total surplus is maximized because the paid subsidy is lower under consumer surplus maximization. The monopsonistic power of the mechanism designer leads to an increase of expected consumer surplus and a reduction of expected firm profits. The imposition of the price constraint, makes the consumer surplus maximizing price mechanism achieve 'third-best' efficient results.

Quantity mechanism. Analogously, I derive functional forms for the quantity mechanism, using Proposition 3.5 and Remark 3.3.

Corollary 3.3. *Suppose that Assumptions 3.1 and 3.2 hold. Suppose further that the regulator uses a quantity mechanism regulate externalities.*

i) The externality reduction under consumer surplus maximization and total surplus maximization are, respectively

$$\hat{Q}^{**} = \left[\tfrac{1}{2}(f^L\theta^L + f^H\theta^H)\left(\tfrac{p^L}{\delta^L + h(\delta^L)} + \tfrac{p^H}{\delta^H + h(\delta^H)}\right)\right]^{\frac{2}{3}} ,$$

and

$$\hat{Q}^{*} = \left[\tfrac{1}{2}(f^L\theta^L + f^H\theta^H)\left(\tfrac{p^L}{\delta^L} + \tfrac{p^H}{\delta^H}\right)\right]^{\frac{2}{3}} .$$

ii) If U is maximized, the expected consumer surplus is

$$\begin{aligned} U(\theta,\delta,\hat{Q}^{**}) =& (f^L\theta^L + f^H\theta^H)\left(\hat{Q}^{**}\right)^{\frac{1}{2}} \\ & - \frac{1}{2}\left(p^L(\delta^L + h(\delta^L)) + p^H(\delta^H + h(\delta^H))\right)\left(\hat{Q}^{**}\right)^2 - \underline{\pi}_j . \end{aligned}$$

iii) If S is maximized, the expected total surplus is

$$S(\theta,\delta,\hat{Q}^{*}) = (f^L\theta^L + f^H\theta^H)\left(\hat{Q}^{*}\right)^{\frac{1}{2}} - \tfrac{1}{2}\left(p^L(\delta^L) + p^H(\delta^H)\right)\left(\hat{Q}^{*}\right)^2 .$$

As firms information rents are no longer part of the surplus measure and as these information rents render emission reduction costly for consumers, the emission reduction under expected consumer surplus maximization is distorted downwards in comparison to expected total surplus maximization. If a quantity constraint is introduced, the expected consumer surplus falls short compared to the expected consumer surplus that can be generated under the unconstrained mechanism. The difference between the expected consumer surplus under the second-best unconstrained mechanism $U(\theta,\delta)$ and the third-best quantity mechanism $U(\theta,\delta,\hat{Q}^{**})$ can hence be interpreted as the efficiency loss from imposing a quantity constraint.

3.5.1 Ex ante comparison

I now turn to the question, which mechanism leads to highest expected total surplus. First, I derive comparative static properties of externality regulation under the three mechanisms, before I clarify conditions under which the price mechanism is better than the quantity mechanism, in the sense that it yields higher expected total surplus. This comparison is in the spirit of Weitzman (1974). He compares price and quantity mechanisms under uncertainty. Contrary to Weitzman, I assume that uncertainty stems from private information, whereas he assumes that uncertainty is driven from incomplete information.[10] I analyze the effects of changing the technology parameter and the pref-

[10]For a comparison of price and quantity mechanisms under incomplete information, see Chapter 2.7.

erence parameter, on the comparative advantage measure. Further, I clarify how the analysis changes if expected consumer surplus is maximized, opposed to total surplus.

Total surplus maximization. The following Proposition derives how the three mechanisms are affected by changes in expected preferences and costs for emission reduction.

Proposition 3.6. *Suppose that Assumptions 3.1 and 3.2 hold.*

i) $\mathbb{E}\left[Q(\theta,\delta)\right] \neq \mathbb{E}\left[Q(\theta,\delta,s^*)\right] = \hat{Q}^*$.

ii) $S(\theta,\delta) > S(\theta,\delta,s^*)$ *and* $S(\theta,\delta) > S(\theta,\delta,\hat{Q}^*)$.

iii) An increase of f^H *has the following implications*

$$\frac{\partial S(\theta,\delta)}{\partial f^H} > 0\ , \qquad \frac{\partial S(\theta,\delta,s^*)}{\partial f^H} > 0 \quad \textit{and} \quad \frac{\partial S(\theta,\delta,\hat{Q}^*)}{\partial f^H} > 0\ .$$

iv) An increase of p^L *has the following implications*

$$\frac{\partial S(\theta,\delta)}{\partial p^L} > 0\ , \qquad \frac{\partial S(\theta,\delta,s^*)}{\partial p^L} >< 0 \quad \textit{and} \quad \frac{\partial S(\theta,\delta,\hat{Q}^*)}{\partial p^L} >< 0\ .$$

The Proposition has several implications that are worth to mention: First, the expected total emission reduction under the price mechanism and the quantity mechanism are equivalent. Since both mechanism account for consumer i's expected preferences and firm j's costs ex ante, the expected reduction levels coincide. However, ex post, less emissions are reduced under the price mechanism than under the quantity mechanism if firm j has high costs ($\delta = \delta^H$) and accordingly, more emissions are reduced when firm j has low costs. The total reduction level does, however, not coincide with the expected emission reduction under the unconstrained mechanism.[11]

Second, there is a measure for how severe the efficiency loss is under the constrained price and quantity mechanisms, compared to the unconstrained mechanism. I can evaluate the impact of price and quantity constraints by comparing the difference between expected total surplus under price and quantity mechanism with the expected surplus under the unconstrained unconstrained mechanism.

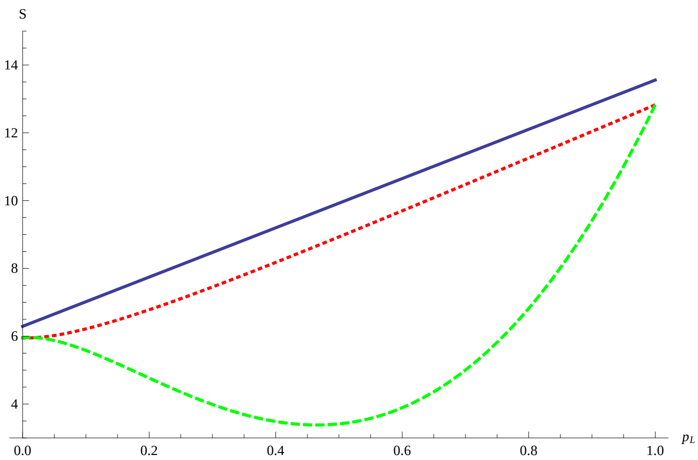
Third, I can analyze how an increase of expected preferences and expected costs affects the expected total surplus under the three mechanisms. If one focuses on an increase of consumer's expected preference, this leads to higher expected total surplus under all three mechanisms. If a reduction in expected costs is considered (p^L increases), the

[11] Whenever $\left(f^L(\theta^L)^{\frac{2}{3}} + f^H(\theta^H)^{\frac{2}{3}}\right)\left(p^L\left(\frac{1}{\delta^L}\right)^{\frac{2}{3}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{2}{3}}\right) > \hat{Q}^* = \mathbb{E}\left[Q(\theta,\delta,s^*)\right]$, the unconstrained mechanism leads to more emission reduction ex ante than the price and the quantity mechanism.

expected total surplus is affected differently under the three mechanisms. Under the unconstrained mechanism a decrease in expected costs leads to higher expected total surplus. By contrast, if a quantity mechanism is used, expected total surplus is lower when the cost parameters are uniformly distributed, compared to the case where the probabilities are distributed unevenly between the two possible realization of the cost parameter. If firm j has high costs, ex post, it needs to reduce more emissions than the efficient amount. Contrary, if firm j has low costs, it reduces less than the efficient amount. As costs of emission reduction are convex, this misspecification is more costly when cost parameters are equally distributed than when they are distributed unequally. The impact of an increase of p^L on expected total surplus when a price mechanism is used is as follows: If p^L is increased, the expected amount of emission reduction is increased and the expected costs of emission reduction are increased. If p^L is marginally increased from 0, then the second effect dominated the first under some parameter constellations. If p^L is increased even further, the first effect dominates the latter.

The following Figure depicts a numerical example for how a change in p^L affects the expected total surplus.

Figure 3.1: Expected total surplus under the three mechanisms



The parameter choices are: $\theta^L = 5$, $\theta^H = 15$, $f^L = f^H = 0.5$, $\delta^L = 1$ and $\delta^H = 10$. The blue linear curve is the expected surplus when the unconstrained mechanism is used, the red dotted curve is the expected surplus when a price mechanism is used and the green dashed curve depicts the expected surplus when a quantity mechanism is used.

Figure 3.1 shows that the unconstrained mechanism leads to higher expected total surplus than the price and quantity mechanism, for all possible values of p^L. When $p^L = 0$ or when $p^H = 1 - p^L = 0$ (so that there is certainty on firm j's expected total costs), the price mechanism and the quantity mechanism lead to the same expected total surplus. As there is still uncertainty about consumer i's preferences for emission reduction, price and quantity mechanisms lead to less expected total surplus than the unconstrained mechanism. For the given numerical values, an inspection of Figure 3.1 reveals that under the given functional forms and the given numerical values, the price mechanism leads to higher expected total surplus than the quantity mechanism.

I now turn to the question how the variance of the preference and cost variable affects the expected surplus under the three mechanisms. The following Figures illustrate a change in the variance for a numerical example.

Figure 3.2: Changing the variance of the preference parameter

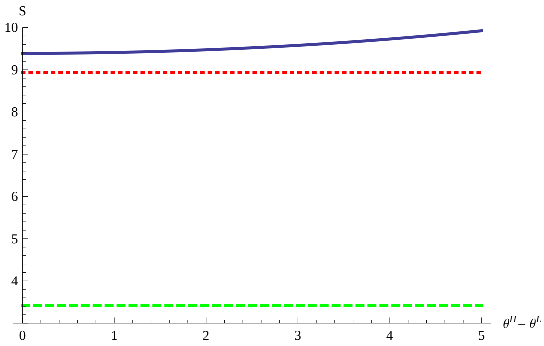
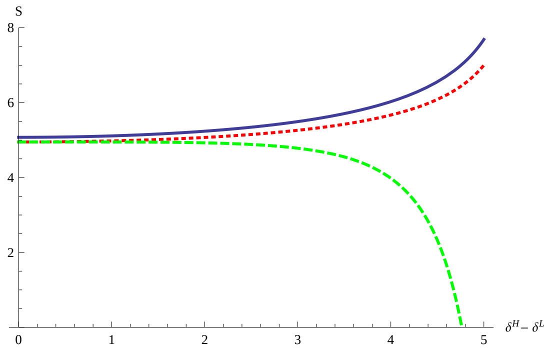


Figure 3.3: Changing the variance of the cost parameter



Numerical examples: The parameter choices are $f^L = f^H = p^L = p^H = 0.5$. In Figure 3.2, on the left, $\delta^L = 5$ and $\delta^H = 15$ and $\mathbb{E}[\theta] = 10$. In Figure 3.3, on the right, $\theta^L = 5$, $\theta^H = 15$ and $\mathbb{E}[\delta] = 5.5$. The blue solid curve is the expected total surplus function when the unconstrained mechanism is used, the red dotted curve is the expected total surplus function when a price mechanism is used and the green dashed curve depicts the expected total surplus function when a quantity mechanism is used.

A change in the variance of preferences does not have an influence on expected surplus when either the price mechanism or the quantity mechanism is used by the regulator, as illustrated in Figure 3.2. The reason is that the preferences enter in such way into the expected surplus function that only the expected value but not the variance affects surplus (see Corollaries 3.2 iii) and 3.3 iii)). The expected surplus under the unconstrained mechanism increases in the difference of θ^H and θ^L, keeping the expected preferences constant. The preference parameters enter the expected surplus function in a convex way, so that the higher the variance of θ, the higher is the expected surplus.

Figure 3.3 illustrates that an increase in the variance of the cost parameter affects price and quantity mechanisms in different ways. The higher the variance of the cost parameter, the lower is the expected surplus under the quantity mechanism. As the quantity level is fixed, the emission reduction is very costly if firm j has high costs ex post. Due to the convexity of the cost function this increase in costs is more severe than the lower costs in case where firm j has low costs. The price mechanism allows for an adaption of the quantity, so that an increase in the variance does lead to an increase in the expected total surplus.

Prices versus quantities. I now turn to the question whether the price mechanism or the quantity mechanism leads to higher expected total surplus. I define the difference between expected total surplus between the price mechanism and the quantity mechanism as

$$\gamma_S = S(\theta, \delta, s^*) - S(\theta, \delta, \hat{Q}^*) \ .$$

Corollary 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then* $\gamma_S > 0$ *if*

$$4 \left(\frac{p^L \left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H \left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - 1 \right) > \left(1 - (p^L \delta^L + p^H \delta^H) \left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right) \ .$$

Corollary 3.4 provides a condition for the price mechanism to lead to higher expected total surplus than the quantity mechanism. As the price mechanism leads in expectation to the same emission reduction than the quantity mechanism, the ex ante benefits of emission reduction are equal; hence, consumer i's preferences do not influence the comparative advantage measure γ_S. From Proposition 3.4 it is know that $\gamma_S = 0$ for $p^L = 0$ and for $p^L = 1$. For the given functional forms of Assumptions 3.1 and 3.2, the marginal benefit curve is relatively flat compared to the marginal cost curve. It is therefore not surprising that the price mechanism leads to higher expected surplus than the quantity mechanism, for all numerical values in Figure 3.1.

Consumer surplus maximization. I now turn to the question how the mechanism respond to changes in parameters when consumer surplus is maximized.

Proposition 3.7. *Suppose that Assumptions 3.1 and 3.2 hold. The regulator wants to maximize* U*.*

i) $\mathbb{E}\left[Q^{**}(\theta, \delta)\right] \neq \mathbb{E}\left[Q(\theta, \delta, s^{**})\right] \neq \hat{Q}^{**}$.

ii) $U(\theta, \delta)^{**} > U(\theta, \delta, s^{**})$ *and* $U(\theta, \delta)^{**} > U(\theta, \delta, \hat{Q}^{**})$.

iii) An increase of f^H *has the following implications*

$$\frac{\partial U(\theta, \delta)}{\partial f^H} > 0 \ , \qquad \frac{\partial U(\theta, \delta, s^*)}{\partial f^H} > 0 \quad \text{and} \quad \frac{\partial U(\theta, \delta, \hat{Q}^*)}{\partial f^H} > 0 \ .$$

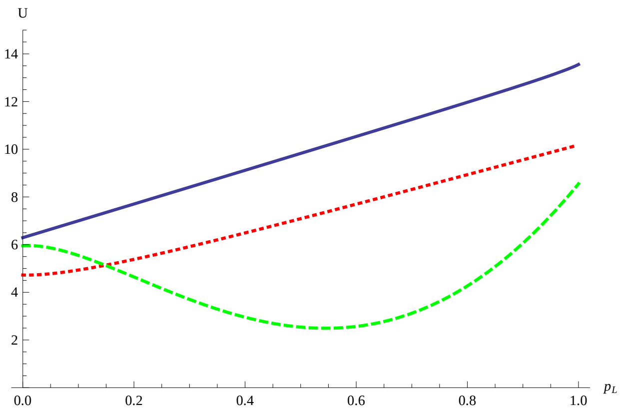
iv) An increase of p^L *has the following implications*

$$\frac{\partial U(\theta, \delta)}{\partial p^L} > 0 \ , \qquad \frac{\partial U(\theta, \delta, s^*)}{\partial p^L} >< 0 \quad \text{and} \quad \frac{\partial U(\theta, \delta, \hat{Q}^*)}{\partial p^L} >< 0 \ .$$

When expected consumer surplus is maximized, the emission reduction under the quantity mechanism and the price mechanism do no longer coincide. When the mechanism designer uses a quantity mechanism, virtual costs of firms are equal. Contrary, under a price mechanism, marginal costs are equal. When consumer surplus is maximized, both mechanisms account differently for firms private information, so that the expected emission reduction differs.

As in Proposition 3.6, the unconstrained mechanism is more efficient than the price and the quantity mechanism. Further, an increase of f^H leads to an increase of expected consumer surplus under all three mechanisms. And again, an increase of p^L leads to an increase of expected consumer surplus under the unconstrained mechanism, whereas under the price and quantity mechanism, this relation is not linear, as the following Figure illustrates.

Figure 3.4: Expected consumer surplus under the three mechanisms



The parameter choices are: $\theta^L = 5$, $\theta^H = 15$, $f^L = f^H = 0.5$, $\delta^L = 1$ and $\delta^H = 10$. The blue continuous curve is the expected surplus when the unconstrained mechanism is used, the red dotted curve is the expected surplus when a price mechanism is used and the green dashed curve depicts the expected surplus when a quantity mechanism is used.

Prices versus quantities. I now turn to the question whether the price or the quantity mechanism leads to higher expected consumer surplus. Similar to the case above, I define the difference between expected consumer surplus between the price mechanism and the quantity mechanism as

$$\gamma_U = U(\theta, \delta, s^{**}) - U(\theta, \delta, \hat{Q}^{**}) \ .$$

Corollary 3.5. *Suppose that Assumptions 3.1 and 3.2 hold.* $\gamma_U > 0$ *if*

$$\frac{1}{4}^{\frac{1}{3}} \left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}} \left(\frac{p^L \left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H \left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - \frac{1}{4}^{4} \right) >$$

$$\left(\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}} \left(\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{7}{3}} \frac{\left(p^L(\delta^L + h(\delta^L)) + p^H(\delta^H + h(\delta^H))\right)}{\left(\frac{p^L}{\delta^L+h(\delta^L)} + \frac{p^H}{\delta^H+h(\delta^H)}\right)^{-1}} \right) .$$

The Corollary establishes under which parameter constellations the price mechanism leads to higher expected consumer surplus than the quantity mechanism. The distribution of firm j's technology is decisive to evaluate whether the price or the quantity mechanisms leads to higher expected consumer surplus because it affects not only the expected reduction costs but also expected information rents under the quantity mechanism. If high information rents need to be paid when a quantity mechanism is used, in comparison to subsidy payments, the price mechanism is likely to lead to higher expected consumer surplus than the quantity mechanism. The Corollary shows that when there is certainty on the costs ($p^L = 0$ or $p^L = 1$), the comparative advantage is different from zero ($\gamma_U \neq 0$).

Figure 3.5 and 3.6 illustrate the implications of a change in p^L on the comparative advantage measure when either expected consumer surplus is maximized or expected total surplus is maximized. The numerical parameter values are identical for both cases.

Figure 3.5: Comparative advantage under total surplus maximization

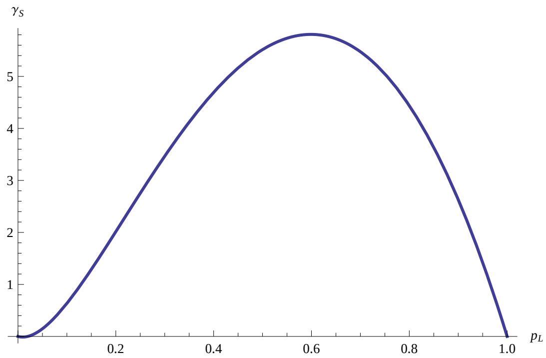
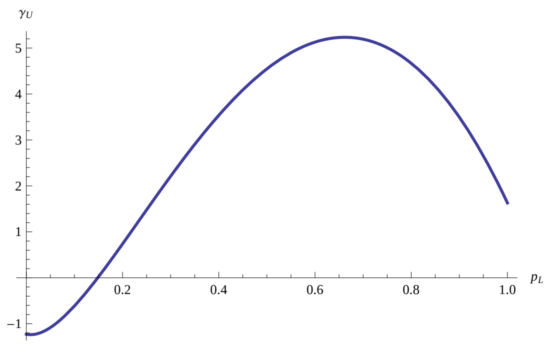


Figure 3.6: Comparative advantage under consumer surplus maximization



The parameter choices are: $\theta^L = 5$, $\theta^H = 15$, $\delta^L = 1$, $\delta^H = 10$ and $f^L = f^H = 0.5$. The figures show that for the given functional forms and numerical values, the price mechanism leads to higher expected total surplus than the quantity mechanism. Under expected consumer surplus maximization, however, the quantity mechanism is better if the probability on the low cost firm is low.

When expected total surplus is maximized, the price mechanism is better than the quantity mechanism. This unambiguousness vanishes if instead expected consumer surplus is maximized. Here, for low values of p^L, the quantity mechanism leads to higher expected consumer surplus than the price mechanism. Hence, if more weight is put on consumer surplus, having certainty on the amount of emission reduction becomes more important and the price mechanism leads to higher surplus for less situations.

3.5.2 Ex post comparison

Price and quantity mechanisms not only differ under the employed surplus measure but also whether the comparison takes place ex ante or ex post. I denote in the following the ex post realization of consumer i's preference parameter with $\hat{\theta}$, and the ex post realization of firm j's cost parameter with $\hat{\delta}$.

Suppose that Assumptions 3.1 and 3.2 hold. Under the price mechanism, ex post total surplus is

$$s(\hat{\theta}, \hat{\delta}, s^*) = \hat{\theta} \left(\frac{s^*}{\hat{\delta}} \right)^{\frac{1}{2}} - \frac{1}{2} \frac{s^{*2}}{\hat{\delta}} \ .$$

Under the quantity mechanism, ex post total surplus is

$$s(\hat{\theta}, \hat{\delta}, \hat{Q}^*) = \hat{\theta} (\hat{Q}^*)^{\frac{1}{2}} - \frac{1}{2} \left(\hat{\delta} \right) (\hat{Q}^*)^2 \ .$$

I define the ex post difference in total surplus as γ_S^P, so that

$$\gamma_S^P = s(\hat{\theta}, \hat{\delta}, s^*) - s(\hat{\theta}, \hat{\delta}, Q^*) \ .$$

Whenever γ_S^P is positive, the price mechanism leads to higher total surplus than the quantity mechanism.

Example: Suppose that $\theta^L = 5$, $\theta^H = 12$, $\delta^L = 1$ and $\delta^H = 3$. Also assume that $f^L = f^H = p^L = p^H = 0.5$. For these numerical values, the unconstrained quantity mechanism fixes an emission level $\hat{Q}^* = 2.00$, the unconstrained price mechanism fixes a subsidy $s^* = 3.00$. When the price mechanism is used, total surplus is given by

$$s(\theta^L, \delta^L, s^*) = 4.15, \quad s(\theta^H, \delta^L, s^*) = 16.29, \quad s(\theta^L, \delta^H, s^*) = 3.49, \quad s(\theta^H, \delta^H, s^*) = 10.50 \ .$$

When the regulator uses the quantity mechanism, total surplus is given by

$$s(\theta^L, \delta^L, \hat{Q}^*) = 5.07, \ s(\theta^H, \delta^L, \hat{Q}^*) = 14.98, \ s(\theta^L, \delta^H, \hat{Q}^*) = 1.06, \ s(\theta^H, \delta^H, \hat{Q}^*) = 10.97 \ .$$

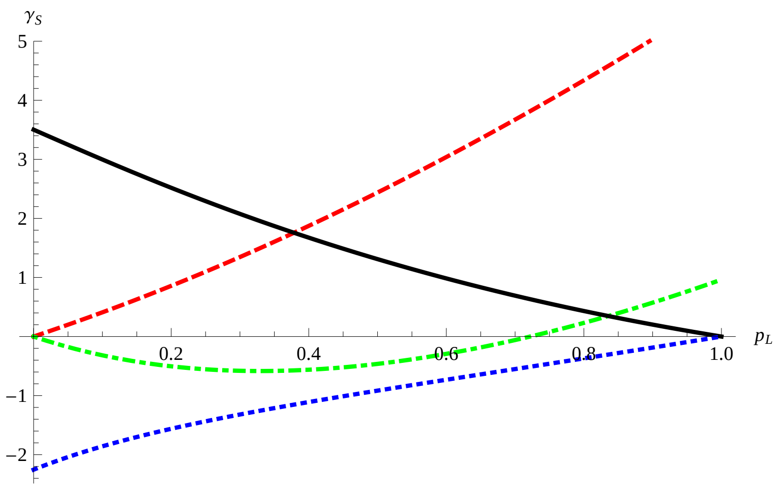
This leads to a comparative advantage of

$$\gamma_S^P(\theta^L, \delta^L) = -0.92, \quad \gamma_S^P(\theta^H, \delta^L) = 1.31, \quad \gamma_S^P(\theta^L, \delta^H) = 2.43, \quad \gamma_S^P(\theta^H, \delta^H) = -0.47 \ .$$

Ex ante, the price mechanism leads to higher expected total surplus than the quantity mechanism.[12] Nevertheless, ex post, the quantity mechanism can lead to higher total surplus than the price mechanism. An inspection of these numbers reveals that there are two states of the economy, (θ^L, δ^L) and (θ^H, δ^H), where the quantity mechanism leads to higher total surplus than the price mechanism, for $p^L = 0.5$. For the two other states, (θ^H, δ^L) and (θ^H, δ^H), the price mechanism leads to higher total surplus.

The comparative advantage is, among other things, driven by the distribution of firm j's cost parameter. The following Figure illustrates the impact of a change in the distribution of costs on the comparative advantage measure. I focus on a change in p^L because this change influences price and quantity mechanisms in different ways. In particular, it demonstrates that the comparative advantage measure, for the four states, is differently affected by a change of the technology distribution. Based on the numerical values, for all values of p^L, there is at least one state of the economy where the quantity mechanism leads to higher ex post surplus, although the price mechanism leads to higher expected surplus ex ante.

Figure 3.7: Comparative advantage – ex post



The parameter choices are as in the Example above. The black decreasing curve, is the comparative advantage function for the state where the consumer has a high valuation for emission reduction and the firm has low costs, $\gamma_S^P(\theta^H, \delta^L)$. The blue dotted curve is the comparative advantage measure for the state $\gamma_S^P(\theta^L, \delta^L)$, and $\gamma_S^P(\theta^L, \delta^H)$ is represented by the red dashed function. The green dashed-dotted function represents $\gamma_S^P(\theta^H, \delta^H)$.

[12] $\gamma_S = \frac{1}{4}\left(\gamma_S^P(\theta^L, \delta^L) + \gamma_S^P(\theta^L, \delta^H) + \gamma_S^P(\theta^H, \delta^L) + \gamma_S^P(\theta^H, \delta^H)\right) > 0.$

3.6 Concluding remarks

This chapter has provided a way to deal with asymmetric information in models of externality regulation, in which the restriction to price and quantity mechanisms makes it impossible to reach efficient outcomes. Looking at the case of emission reduction where firms have private information about their technology, conditions have been shown to see whether the price or the quantity mechanism leads to higher efficiency. Different surplus notions have been compared. The analysis has established a link between the surplus measure and the level of emission reduction. An interesting insight of this analysis was that the regulator should reduce less emissions if consumer surplus had been maximized, compared to the case where total surplus maximization was the regulator's target. I have argued that this result is independent from the mechanism being used.

I have demonstrated that distributive considerations have a great significance for the regulator's choice between price and quantity mechanisms. This observation suggests that when regulatory agencies decide on externality regulation, they first need to address, which distributive goals they want to achieve, before comparing both mechanisms with respect to efficiency.

The analysis has abstracted from the origin of the externality and assumed that only firms respond to regulation. Another important extension of my research would study a firm sector that not only reduces externalities but also engages in a profit maximizing activity. This would require a good understanding of how different sectors can deal with higher environmental standards. Further, the model could be extended so that consumers can as well reduce externalities. This considerations are of obvious importance. However, I think that my framework is suited to combine the independent private values model of mechanism design with externality regulation by means of price and quantity mechanisms.

Appendix 3.A Proofs of Propositions

Proof of Proposition 3.2.

The regulator needs to maximize the consumer surplus subject to the condition that consumers transfers need to cover the resource requirement.

$$max \quad \mathbb{E}\left[\sum_{i=1}^{n}(\theta_i b\left(Q(\theta,\delta)\right) - t_i(\theta,\delta))\right] ,$$

subject to a budget constraint

$$\mathbb{E}\left[\sum_{i=1}^{n} t_i(\theta,\delta)\right] \geq R(\theta,\delta) .$$

The budget constraint is fulfilled with equality, as otherwise the aggregated expected consumer surplus could be increased by reducing consumers' transfers, without violating the constraint. Using the result of Proposition 3.1, the regulator's objective function can be written as

$$max \quad \mathbb{E}\left[\sum_{i=1}^{n}(\theta_i b\left(\sum_{j=1}^{m} q_j(\theta,\delta)\right)\right] - \sum_{j=1}^{m}\underline{\pi}_j - \mathbb{E}\left[\sum_{j=1}^{m}\left(\delta_j + h\left(\delta_j\right)\right) c\left(q_j(\theta,\delta)\right)\right] .$$

If the surplus function is maximized with respect to $qj(\theta,\delta)$, this yields the following first order condition, for all firms $j = \{1, ..., m\}$

$$\sum_{i=1}^{n}\theta_i b'\left(\sum_{j=1}^{m} q_j^{**}(\theta,\delta)\right) = \left(\delta_j + h(\delta_j)\right) c'\left(q_j^{**}(\theta,\delta)\right) .$$

Proof of Proposition 3.3.

The regulator needs to maximize the consumer surplus subject to the condition that total costs need to be covered .

$$max \quad \mathbb{E}\left[\sum_{i=1}^{n}(\theta_i b\left(Q(\theta,\delta,s)\right) - t_i(\theta,\delta,s))\right] ,$$

subject to a budget constraint

$$\mathbb{E}\left[\sum_{i=1}^{n} t_i(\theta,\delta,s)\right] \geq R(\theta,\delta,s) .$$

The budget condition needs to hold with equality, as otherwise the aggregated expected consumer surplus could be maximized by reducing the consumers' transfers without violating the constraint. The surplus function can therefore be simplified such that

$$max \quad \mathbb{E}\left[\sum_{i=1}^{n}\theta_i b(Q(\theta,\delta,s)) - sQ(\theta,\delta,s)\right] - \sum_{j=1}^{m}\underline{\pi}_j \; .$$

If the surplus function is maximized with respect to s, the following first order condition needs to hold

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i b'\left(Q(\theta,\delta,s^{**})\right)\frac{\partial Q(\theta,\delta,s^{**})}{\partial s}^{**}\right] = \mathbb{E}[Q(\theta,\delta,s^{**})] + s^{**}\mathbb{E}\left[\frac{\partial Q(\theta,\delta,s^{**})}{\partial s^{**}}\right] \; .$$

Proof of Proposition 3.4.

ii) Suppose there is some exogenously given amount $R(\delta,\hat{Q})$ coming from the regulator in order to finance the externality reduction, $R(\delta,\hat{Q}) \geq \mathbb{E}\left[\sum_{j=1}^{m} r_j(\delta,\hat{Q})\right]$.
The objective is to choose $q_j : \delta \times \hat{Q} \mapsto q_j(\delta,\hat{Q})$ and $r_j : \delta \times \hat{Q} \mapsto r_j(\delta,\hat{Q})$ so as to minimize

$$min \quad \mathbb{E}\left[\sum_{j=1}^{m} r_j(\delta,\hat{Q})\right] = \sum_{l=1}^{r} p^l R_j(\delta^l) \; ,$$

subject to the constraints (IC_F), (PC_F) and (QU).
I work with interim participation constraints; these depend on the parameter $\underline{\pi}_j$. I will argue later that this model is flexible enough to cover the case of ex ante participation constraints, which will be satisfied if the parameter $\underline{\pi}_j$ takes appropriate values.
Step 1: I treat the provision rule $(q_j)_{j\in J}$ as exogenously given and solve the cost minimizing transfer scheme that implements this provision rule.
By Proposition 3.1, I can derive

$$\begin{aligned} R(\delta,\hat{Q}) &= \sum_{l=1}^{r} p^l R_j(\delta^l) \\ &= \sum_{j=1}^{m}\underline{\pi}_j + \mathbb{E}\left[\sum_{j=1}^{m}\left(\delta_j + h(\delta_j)\right) c\left(q_j\left(\delta,\hat{Q}^{**}\right)\right)\right] \; . \end{aligned}$$

Step 2: To complete the proof, I need to show that the solution to the cost minimizing problem in *Step 1* satisfies all the participation constraints in (PC_F) and all locally-downward binding incentive compatibility conditions.
Suppose that emission reduction is such that the following monotonicity constraint holds: for all l, $C_j(\delta^l) \geq C_j(\delta^{l+1})$. I show that under this assumption there is a mecha-

nism such that $R(\delta,\hat{Q}) = \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c_j\left(q_j(\delta,\hat{Q}^{**})\right)\right] + \sum_{j=1}^{m}\underline{\pi}_j$, satisfies all the participation constraints in (PC_F) and all incentive compatibility conditions.
With the help of Lemma 1.11 in Chapter 1, I find that if the firms' participation constraints holds for $\delta = \delta^r$, then it holds for all $\delta \neq \delta^r$.

Justification of interim participation constraints.
To justify interim participation constraints, I want to state the following on the ex ante participation constraint: Under the cost-minimizing mechanism the expected interim payoff of a lowest technology firm equals $\underline{\pi}$ and the expected interim payoffs of firms with better technology equal (due to the binding incentive compatibility constraint)

$$R_j(\delta^l) - \delta^l C_j(\delta^l) = \underline{\pi}_j + \frac{P(\delta^l)}{p(\delta^l)} C_j(\delta^l) \ .$$

Ex ante expected payoffs are

$$\sum_{j=1}^{m} \underline{\pi}_j + \sum_{l=2}^{r} P(\delta^l) C_j(\delta^j) = \sum_{j=1}^{m} \underline{\pi}_j + \mathbb{E}\left[\sum_{j=1}^{m} h\left(\delta_j\right) c\left(q_j\left(\delta,\hat{Q}\right)\right)\right] \ .$$

Ex ante participation constraints are satisfied if

$$\sum_{j=1}^{m} \underline{\pi}_j + \mathbb{E}\left[\sum_{j=1}^{m} h(\delta_j) c\left(q_j\left(\delta,\hat{Q}\right)\right)\right] \geq \underline{\pi}_j \ ,$$

where $\underline{\pi}_j$ is the minimal ex ante payoff. This constraint ensures that the provision of the emission reduction by the firms leads to an Pareto-improvement if considered behind a 'veil of ignorance' where firms can form an expectation about how costly the provision will be, but they are not yet fully informed about their technologies.
iii) As I have shown, the resource requirement of the regulator is

$$R(\delta,\hat{Q}) = \sum_{j=1}^{m} \underline{\pi}_j + \mathbb{E}\left[\sum_{j=1}^{m} \left(\delta_j + h(\delta_j)\right) c\left(q_j\left(\delta,\hat{Q}\right)\right)\right] \ ,$$

whereas the costs that the production sector face are equal to

$$K(\delta,\hat{Q}) = \mathbb{E}\left[\sum_{j=1}^{m} \delta_j c\left(q_j\left(\delta,\hat{Q}\right)\right)\right] \ .$$

The difference is the production sectors' expected payoffs

$$\Pi(\delta,\hat{Q}) = R(\delta,\hat{Q}) - K(\delta,\hat{Q}) = \sum_{j=1}^{m} \underline{\pi}_j + \mathbb{E}\left[\sum_{j=1}^{m} h(\delta_j) c\left(q_j\left(\delta,\hat{Q}\right)\right)\right] .$$

i) It remains to be shown that marginal virtual costs are equalized. The regulator needs to decide how much of the emission reduction is done by firms that differ in their technology. Therefore, the mechanism designer needs to look at firm j's virtual marginal costs. The intersection of the virtual marginal transfer curve for firm j with technology δ^l with the expected marginal benefit curve $\left(\mathbb{E}[MB(Q)]\right)$ of total emission reduction leads to amount $q_j(\delta^l)$ firm j needs to reduce. The emission reduction $q_j(\delta^l)$ needs to satisfy the following condition, for all l

$$\mathbb{E}\left[MB(Q)\right] = \mathbb{E}\left[(\delta_l + h(\delta_l)) \frac{\partial c(q_j(\delta,\hat{Q}))}{\partial q_j(\delta,\hat{Q}))} \frac{\partial q_j(\delta,\hat{Q}))}{\partial \hat{Q}}\right] .$$

For ease of notation, I write: $c'(q_j(\delta,\hat{Q})) := \frac{\partial c(q_j(\delta,\hat{Q}))}{\partial q_j(\delta,\hat{Q})} \frac{\partial q_j(\delta,\hat{Q})}{\partial \hat{Q}}$.
When the regulator decides on the emission reduction of each firm, the marginal benefit curve stays fixed and hence, the virtual marginal cost curves all face the same marginal benefit curve. Therefore, virtual marginal costs need to be equalized, for all δ, and for all firms j and all j'

$$(\delta_j + h(\delta_j))\, c'\left(q_j\left(\delta,\hat{Q}\right)\right) = (\delta_{j'} + h(\delta_{j'}))\, c'\left(q_{j'}\left(\delta,\hat{Q}\right)\right) .$$

Proof of Proposition 3.5.

As consumers' transfers are used to pay firms' transfers, the surplus maximization problem can be stated as follows
Choose a provision rule $Q(\delta\hat{Q}) = \sum_{j=1}^{m} q_j(\delta,\hat{Q})$ in order to maximize

$$\begin{aligned} U(\theta,\delta,\hat{Q}) &= \mathbb{E}\left[\left(\sum_{i=1}^{n}\theta_i\right)\right] b\left(\sum_{j=1}^{m} q_j(\delta,\hat{Q})\right) - \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\delta,\hat{Q}))\right] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^{n}\theta_i\right) b(\hat{Q})\right] - \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c(q_j(\delta,\hat{Q}))\right] . \end{aligned}$$

The solution to this problem is such that, for every $(\delta,\hat{Q})$, the following first order condition is satisfied

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i\right] b'(\hat{Q}^{**}) = \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))\frac{\partial c(q_j(\delta,\hat{Q}))}{\partial q_j(\delta,\hat{Q})}\sum_{j=1}^{m}\frac{\partial q_j(\delta,\hat{Q})}{\partial \hat{Q}^{**}(\delta,\hat{Q})}\right]$$

$$\mathbb{E}\left[\sum_{i=1}^{n}\theta_i\right] b'(\hat{Q}^{**}(\delta,\hat{Q})) = \mathbb{E}\left[\sum_{j=1}^{m}(\delta_j + h(\delta_j))c'(q_j(\delta,\hat{Q}))\right] .$$

The first order condition implies that, for every $\theta, \hat{Q}^{**}(\delta^k,\hat{Q}) \geq Q^{**}(\delta^l,\hat{Q})$, for $k \neq l$. This implies that the monotonicity constraint $C_j(\delta^l) \geq C_j(\delta^{l+1})$, for all l, is satisfied.

Proof of Proposition 3.6

i) When subsidy s^* is introduced, then every firm j reduces $q_j = \frac{s^*}{\delta_j}$. Making use of the expression derived in Corollary 3.2 i), the expected amount of emission reduction under the price mechanism is

$$\mathbb{E}[Q(\delta, s^*)] = \left[(f^L\theta^L + f^H\theta^H)\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right]^{\frac{2}{3}} = \hat{Q}^* .$$

The expected emission reduction under the price and quantity mechanism coincide. To derive the expected emission reduction under the unconstrained mechanism, the possible states of the economy need to be weighted. Using Corollary 3.1 i), the expected emission reduction under the unconstrained mechanism is

$$\mathbb{E}[Q(\theta,\delta)] = (\tfrac{1}{2})^{1/3}(f^L(\theta^L)^{\frac{2}{3}} + f^H(\theta^H)^{\frac{2}{3}})\left(p^L\left(\tfrac{1}{\delta^L}\right)^{\frac{2}{3}} + p^H\left(\tfrac{1}{\delta^H}\right)^{\frac{2}{3}}\right) .$$

ii) The expected total surplus under the unconstrained mechanism is

$$S(\theta,\delta) = \left[\frac{1}{2}^{1/3} - \frac{1}{2}^{\frac{7}{3}}\right]$$
$$\left(f^Lp^L\left(\frac{(\theta^L)^4}{\delta^L}\right)^{\frac{1}{3}} + f^Hp^L\left(\frac{(\theta^H)^4}{\delta^L}\right)^{\frac{1}{3}} + f^Lp^H\left(\frac{(\theta^L)^4}{\delta^H}\right)^{\frac{1}{3}} + f^Hp^H\left(\frac{(\theta^H)^4}{\delta^H}\right)^{\frac{1}{3}}\right) .$$

The expected total surplus under the price mechanism is

$$S(\theta,\delta,s^*) = \left[\frac{1}{2}^{\frac{1}{3}}\frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - \frac{1}{2}^{\frac{7}{3}}\right]\left(f^L\theta^L + f^H\theta^H\right)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}} .$$

The expressions in the first brackets is bigger under the unconstrained mechanism than under the price mechanism because $\frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} < 1$. Further, the expressions in the latter brackets is bigger for the unconstrained mechanism than for the price mech-

anism.
Under the quantity mechanism, the expected total surplus is

$$S(\theta,\delta,\hat{Q}^*)=\left[\frac{1}{2}^{\frac{1}{3}}-\frac{1}{2}^{\frac{7}{3}}(p^L\delta^L+p^H\delta^H)\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)\right](f^L\theta^L+f^H\theta^H)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{3}}.$$

The expression in the first bracket is bigger under the unconstrained mechanism than under the quantity mechanism as $(p^L\delta^L+p^H\delta^H)\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)>1$. Due to Jansen's inequality the expressions in the latter brackets is lower than under the unconstrained mechanism, so that expected total surplus under the quantity mechanism is lower than under the unconstrained mechanism.
iii)

$$\frac{\partial S(\theta,\delta)}{\partial f^H}=p^L\left(\left(\frac{(\theta^H)^4}{(\delta^L)}\right)^{\frac{1}{3}}-\left(\frac{(\theta^L)^4}{(\delta^L)}\right)^{\frac{1}{3}}\right)+p^H\left(\left(\frac{(\theta^H)^4}{(\delta^H)}\right)^{\frac{1}{3}}-\left(\frac{(\theta^L)^4}{(\delta^H)}\right)^{\frac{1}{3}}\right)>0\,.$$

$$\begin{aligned}\frac{\partial S(\theta,\delta,s^*)}{\partial f^H}&=\left[\frac{1}{2}^{1/3}\frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}}+p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{2}}}-\frac{1}{2}^{\frac{7}{3}}\right]\\&\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{3}}\frac{4}{3}((1-f^H)\theta^L+f^H\theta^H)^{\frac{1}{3}}(\theta^H-\theta^L)>0\,.\end{aligned}$$

$$\begin{aligned}\frac{\partial S(\theta,\delta,\hat{Q}^*)}{\partial f^H}&=\left[\frac{1}{2}^{\frac{1}{3}}-\frac{1}{2}^{\frac{7}{3}}(p^L\delta^L+p^H\delta^H)\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)\right]\\&\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{3}}\frac{4}{3}((1-f^H)\theta^L+f^H\theta^H)^{\frac{1}{3}}(\theta^H-\theta^L)>0\,.\end{aligned}$$

iv)

$$\frac{\partial S(\theta,\delta)}{\partial p^L}=f^L\left(\left(\frac{(\theta^L)^4}{(\delta^L)}\right)^{\frac{1}{3}}-\left(\frac{(\theta^L)^4}{(\delta^H)}\right)^{\frac{1}{3}}\right)+f^H\left(\left(\frac{(\theta^H)^4}{(\delta^L)}\right)^{\frac{1}{3}}-\left(\frac{(\theta^H)^4}{(\delta^H)}\right)^{\frac{1}{3}}\right)>0\,.$$

Price and the quantity mechanism are not monotonically increasing in p^L, see numerical Example in Figure 3.1.

Proof of Proposition 3.7

i) When the subsidy s^{**} is introduced, the expected emission reduction under a price mechanism is $s^{**}\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)$. The emission reduction under the quantity mechanism is higher than under the price mechanism if

$$\left[\frac{1}{2}\left(\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)}\right)\right]^{\frac{2}{3}} \geq \left[\frac{1}{4}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right]^{\frac{2}{3}}$$
$$\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)} \geq \frac{1}{2}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)$$

Using Corollary 3.1 i), the expected emission reduction under the unconstrained consumer surplus maximizing mechanism is

$$\mathbb{E}[Q(\theta,\delta)] = (\tfrac{1}{2})^{\frac{1}{3}}(f^L(\theta^L)^{\frac{2}{3}} + f^H(\theta^H)^{\frac{2}{3}})\left(p^L\left(\tfrac{1}{\delta^L+h(\delta^L)}\right)^{\frac{2}{3}} + p^H\left(\tfrac{1}{\delta^H+h(\delta^H)}\right)^{\frac{2}{3}}\right) .$$

ii) The expected consumer surplus under the unconstrained consumer surplus maximizing mechanism is

$$U(\theta,\delta) = \left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{4}{3}}\right]\left(f^L p^L \left(\frac{(\theta^L)^4}{\delta^L + h(\delta^L)}\right)^{\frac{1}{3}} + f^H p^L \left(\frac{(\theta^H)^4}{\delta^L + h(\delta^L)}\right)^{\frac{1}{3}}\right.$$
$$\left. + f^L p^H \left(\frac{(\theta^L)^4}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}} + f^H p^H \left(\frac{(\theta^H)^4}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}}\right) .$$

The expected consumer surplus under the price mechanism is

$$U(\theta,\delta,s^{**}) = \left[\frac{1}{4}^{\frac{1}{3}} \frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - \frac{1}{4}^{\frac{4}{3}}\right]\left(f^L\theta^L + f^H\theta^H\right)^{4/3}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}} .$$

The expressions in the first brackets is bigger under the unconstrained mechanism than under the price mechanism because $\frac{1}{4}^{\frac{1}{3}} \frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - \frac{1}{4}^{\frac{4}{3}} < 1$. Further, the expressions in the latter brackets is bigger for the unconstrained mechanism than for the price mechanism. The expected consumer surplus under the quantity mechanism is

$$U(\theta,\delta,\hat{Q}^{**}) = (f^L\theta^L + f^H\theta^H)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)}\right)^{\frac{1}{3}}$$
$$\left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{4}{3}}\left(p^L(\delta^L + (h(\delta^L))) + p^H(\delta^H + (h(\delta^H)))\right)\left(\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)}\right)\right] .$$

The expression in the first bracket is bigger under the unconstrained mechanism than under the quantity mechanism as $\left(\frac{p^L}{\delta^L+h(\delta^L)} + \frac{p^H}{\delta^H+h(\delta^H)}\right) > 1$. Due to Jansen's inequality the expressions in the latter brackets is lower than under the unconstrained mechanism, so that expected total surplus under the quantity mechanism is lower than under the

unconstrained mechanism.
iii) By Assumption 3.2, $1 - f^H = f^L$. An increase of the expected valuation therefore leads to an increase of expected consumer surplus under all three mechanisms.
iv)

$$\frac{\partial U(\theta,\delta)}{\partial p^L} = f^L \left(\left(\frac{(\theta^L)^4}{(\delta^L + h(\delta^L))} \right)^{\frac{1}{3}} - \left(\frac{(\theta^L)^4}{(\delta^H + h(\delta^H))} \right)^{\frac{1}{3}} \right) + f^H \left(\left(\frac{(\theta^H)^4}{(\delta^L + h(\delta^L))} \right)^{\frac{1}{3}} - \left(\frac{(\theta^H)^4}{(\delta^H + h(\delta^H))} \right)^{\frac{1}{3}} \right) > 0 \,.$$

Price and the quantity mechanism are not monotonically increasing in p^L, see numerical Example in Figure 3.4.

Appendix 3.B Proofs of Corollaries

Proof of Corollary 3.1

i) The optimal emission level under consumer surplus maximization needs to satisfy the condition in Proposition 3.2. Using Assumptions 3.1 and 3.2 gives the following expression

$$Q^{**}(\theta,\delta)=\left(\frac{1}{2}\frac{\theta}{\delta+h(\delta)}\right)^{\frac{2}{3}} .$$

The optimal emission reduction under total surplus maximization needs to satisfy the condition in Remark 3.1. Making use of Assumptions 3.1 and 3.2, yields the following expression

$$Q^{*}(\theta,\delta)=\left(\frac{1}{2}\frac{\theta}{\delta}\right)^{\frac{2}{3}} .$$

ii) Using Assumptions 3.1 and 3.2, together with the expressions derived in part i), lead to the expected consumer surplus stated in the body of the text.
iii) Using Assumptions 3.1 and 3.2, together with the expressions derived in part i), lead to the expected total surplus stated in the body of the text.

Proof of Corollary 3.2.

i) The expected emission reduction is $\mathbb{E}[Q(\delta,s)]=s\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)$. In case of consumer surplus maximization, emission reduction is given by

$$\left(\frac{1}{2}(f^L\theta^L+f^H\theta^H)E[Q(\delta,s^{**})]^{-0.5}-s^{**}\right)\frac{\partial\mathbb{E}[Q(\delta,s^{**})]}{\partial s^{**}}=\mathbb{E}[Q(\delta,s^{**})] .$$

Hence,

$$s^{**}=\frac{\left(\frac{1}{4}(f^L\theta^L+f^H\theta^H)\right)^{\frac{2}{3}}}{\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{3}}} .$$

In case of total surplus maximization, emission reduction is given by

$$\frac{1}{2}(f^L\theta^L+f^H\theta^H)\mathbb{E}[Q(\delta,s^{*})]^{-0.5}=s^{*} .$$

Hence,

$$s^{*}=\frac{\left(\frac{1}{2}(f^L\theta^L+f^H\theta^H)\right)^{\frac{2}{3}}}{\left(\frac{p^L}{\delta^L}+\frac{p^H}{\delta^H}\right)^{\frac{1}{3}}} .$$

ii) The expected emission reduction is given by

$$\mathbb{E}[Q(\delta, s^{**})] = s^{**}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right) .$$

Expected consumer surplus is therefore

$$U(\theta, \delta, s^{**}) = \left(f^L\theta^L + f^H\theta^H\right)\left[p^L\left(\tfrac{s^{**}}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\tfrac{s^{**}}{\delta^H}\right)^{\frac{1}{2}}\right] - s^{**2}\left(\tfrac{p^L}{\delta^L} + \tfrac{p^H}{\delta^H}\right) - \underline{\pi}_j .$$

iii) Expected total emission reduction is

$$\mathbb{E}[Q(\delta, s^*)] = \left(\frac{1}{2}(f^L\theta^L + f^H\theta^H)\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right)^{\frac{2}{3}} .$$

Expected total surplus is given by

$$\begin{aligned} S(\theta, \delta) = \left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{7}{3}}\right]\Bigg[& f^L p^L\left(\frac{(\theta^L)^4}{\delta^L}\right)^{\frac{1}{3}} + f^H p^L\left(\frac{(\theta^H)^4}{\delta^L}\right)^{\frac{1}{3}} \\ & + f^L p^H\left(\frac{(\theta^L)^4}{\delta^H}\right)^{\frac{1}{3}} + f^H p^H\left(\frac{(\theta^H)^4}{\delta^H}\right)^{\frac{1}{3}}\Bigg] . \end{aligned}$$

Proof of Corollary 3.3.

Expected consumer surplus is

$$\begin{aligned} U(\theta, \delta, \hat{Q}) =& (f^L\theta^L + f^H\theta^H)\hat{Q}^{\frac{1}{2}} - \frac{1}{2}\, p^L(\delta^L + h(\delta^L))q(\delta^L, \hat{Q})^2 \\ & - \frac{1}{2}\, p^H(\delta^H + h(\delta^H))q^H(\delta, \hat{Q})^2 - \underline{\pi}_j . \end{aligned}$$

Making use of the fact that expected virtual marginal costs need to be equal, this can be expressed as

$$U(\theta, \delta, \hat{Q}) = (f^L\theta^L + f^H\theta^H)\hat{Q}^{\frac{1}{2}} - \frac{1}{2}\hat{Q}^2 - \underline{\pi}_j .$$

The first order condition is

$$\hat{Q}^{**} = \left[(f^L\theta^L + f^H\theta^H)\left(\frac{p^L}{\delta^L + h(\delta^L)} + \frac{p^H}{\delta^H + h(\delta^H)}\right)\right]^{\frac{2}{3}} .$$

Expected total surplus is

$$S(\theta, \delta, \hat{Q}) = (f^L\theta^L + f^H\theta^H)\hat{Q}^{\frac{1}{2}} - \frac{1}{2}\, p^L\delta^L q(\delta^L, \hat{Q})^2 - \frac{1}{2}\, p^H\delta^H q(\delta^H, \hat{Q})^2 .$$

Making use of the fact that expected marginal costs need to be equal, this can be expressed as

$$S(\theta,\delta,\hat{Q}) = (f^L\theta^L + f^H\theta^H)\hat{Q}^{\frac{1}{2}} - \frac{1}{2}\hat{Q}^2\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right) .$$

The first order condition is

$$\hat{Q}^* = \left[\frac{1}{2}(f^L\theta^L + f^H\theta^H)\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right]^{\frac{2}{3}} .$$

Proof of Corollary 3.4 If the subsidy s^* that maximizes total surplus is plugged in the expression in Corollary 3.2 iii), this yields

$$S(\theta,\delta,s^*) = (f^L\theta^L + f^H\theta^H)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}}\left[\frac{1}{2}^{\frac{1}{3}}\frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{2}}} - \frac{1}{2}^{\frac{7}{3}}\right] .$$

If the subsidy $\hat{Q}^*$ that maximizes total surplus is plugged in the expression in Corollary 3.3 iii), this yields

$$S(\theta,\delta,\hat{Q}^*) = (f^L\theta^L + f^H\theta^H)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}}\left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{7}{3}}(p^L\delta^L + p^H\delta^H)\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)\right] .$$

Comparing these expressions and rearranging terms yield the expression for the comparative advantage measure when total surplus is maximized.

Proof of Corollary 3.5 If the subsidy s^{**} that maximizes total surplus is plugged in the expression in Corollary 3.2 ii), this yields

$$U(\theta,\delta,s^{**}) = (f^L\theta^L + f^H\theta^H)^{\frac{4}{3}}\left[\frac{1}{4}^{\frac{1}{3}}\frac{p^L\left(\frac{1}{\delta^L}\right)^{\frac{1}{2}} + p^H\left(\frac{1}{\delta^H}\right)^{\frac{1}{2}}}{\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{6}}} - \frac{1}{4}^{\frac{4}{3}}\left(\frac{p^L}{\delta^L} + \frac{p^H}{\delta^H}\right)^{\frac{1}{3}}\right] .$$

If the subsidy $\hat{Q}^{**}$ that maximizes total surplus is plugged in the expression in Corollary

3.3 ii) yields

$$U(\theta,\delta,\hat{Q}^{**}) = (f^L\theta^L + f^H\theta^H)^{\frac{4}{3}}\left(\frac{p^L}{\delta^L+h(\delta^L)}+\frac{p^H}{\delta^H+h(\delta^H)}\right)^{\frac{1}{3}}$$
$$\left[\frac{1}{2}^{\frac{1}{3}} - \frac{1}{2}^{\frac{4}{3}}(p^L(\delta^L+h(\delta^L)) + p^H(\delta^H+h(\delta^H))\left(\frac{p^L}{\delta^L+h(\delta^L)}+\frac{p^H}{\delta^H+h(\delta^H)}\right)\right] .$$

Comparing these expressions and rearranging terms yield the expression for the comparative advantage measure when consumer surplus is maximized.

4

Robust Mechanism Design and Social Preferences

4.1 Introduction

Inspired by Wilson (1987), Bergemann and Morris (2005) have provided a formalization of mechanisms that are robust in the sense that they do not rely on a common prior distribution of material payoffs. We add another dimension in which we seek robustness. A mechanism that works well under selfish preferences might fail under social preferences. Indeed, behavioral economics has shown that many agents behave socially. One challenge is, though, that social preferences can differ with respect to their nature and intensity, leading to different kinds of social preference models, including altruism, inequity-aversion, and intentionality (Cooper and Kagel, 2013). Because we want a mechanism to work not only for selfish preferences but also for a large set of social preferences, we introduce the notion of social-preference-robust mechanism: a mechanism must not depend on specific assumptions about the nature and intensity of selfish and social preferences. The following quote of Wilson (1987), which can also be found in Bergemann and Morris (2005), suggests that our approach is a natural next step:

> *"Game theory has a great advantage in explicitly analyzing the consequences of trading rules that presumably are really common knowledge; it is deficient to the extent it assumes other features to be common knowledge, such as one player's probability assessment about another's* ***preferences*** *or* ***information*** *(Emphasis added). I foresee the progress of game theory as depending on suc-*

> *cessive reductions in the base of common knowledge required to conduct useful analyses of practical problems. Only by repeated weakening of common knowledge assumptions will the theory approximate reality."*

While Bergemann and Morris (2005) have focused on common knowledge assumptions regarding the information structure, we seek robustness with respect to common knowledge assumptions on the content of preferences. To this end, we study two classic applications of mechanism design theory, the bilateral-trade problem due to Myerson and Satterthwaite (1983) and a specific version of the optimal income tax problem due to Mirrlees (1971). We show, and provide complementary laboratory evidence, that solutions to these problems, which are derived under the assumption of selfish preferences, are not robust to the possibility that individuals are motivated by social preferences. We then introduce the notion of a social-preference-robust mechanism, and derive mechanisms that are optimal in this class. Finally, we use laboratory experiments to compare the optimal mechanisms under selfish preferences and the optimal social-preference-robust mechanisms.

The bilateral trade problem. The bilateral-trade problem provides us with a simple, and stylized setup that facilitates a clear exposition of our approach. Moreover, it admits interpretations that are of interest in public economics, environmental economics, or contract theory. The basics are as follows: A buyer either has a high or low valuation of a good produced by a seller. The seller either has a high or a low cost of producing the good. An economic outcome specifies, for each possible combination of the buyer's valuation and the seller's cost, the quantity to be exchanged, the price paid by the buyer and the revenue received by the seller. Both the buyer and the seller have private information. Thus, an allocation mechanism has to ensure that the buyer does not understate his valuation so as to get a desired quantity at a lower price. Analogously, the seller has to be incentivized so that she does not exaggerate her cost in order to receive a larger compensation. This environment can be reinterpreted as a problem of voluntary public-goods provision in which one party benefits from larger provision levels, relative to some status quo outcome, and the other party is harmed. By how much the first party benefits and the second party loses is private information. The allocation problem then is to determine the public-goods provision level and how the provision costs should be divided between the two parties. It can also be reinterpreted as a problem to control externalities. One party can invest so as to avoid emissions which harm the other party. The cost of the investment to one party and the benefit of reduced emissions to the other party are private information. In a principal-agent-framework, we may think of one party as benefiting from effort that is exerted by the other party. The size of the benefit and the disutility of effort are, respectively, private information of the principal and the agent.

Our analysis proceeds as follows: We first characterize an optimal direct mechanism for the bilateral trade problem under the standard assumption of selfish preferences, i.e. both, the buyer and the seller, are assumed to maximize their own payoff, respectively, and this is common knowledge. We solve for the mechanism that maximizes the seller's expected profits subject to incentive constraints, participation constraints, and a resource constraint. We work with *ex post* incentive and participation constraints, i.e. we insist that after the outcome of the mechanism and the other party's private information have become known, no party regrets to have participated and to have revealed its own information.

Our reason for imposing these constraints is twofold: First, as has been shown by Bergemann and Morris (2005), they imply that a mechanism is robust in the sense that its outcome does not depend on the individual's probabilistic beliefs about the other party's private information. Second, we use the arguments in Bergemann and Morris (2005) for our experimental testing strategy. In their characterization of robust mechanisms *complete information environments* play a key role. In such an environment, the buyer knows the seller's cost and the seller knows the buyer's valuation, and, moreover, this is commonly known among them. The mechanism designer, however, lacks this information and therefore still has to provide incentives for a revelation of privately held information. Bergemann and Morris provide conditions so that the requirement of robustness is equivalent to the requirement that a mechanism generates the intended outcome in every complete information environment, which in turn is equivalent to the requirement that incentive and participation constraints hold in an *ex post* sense.[1]

In our laboratory approach, we investigate the performance of an optimally designed robust mechanism in all complete information environments. This approach is useful because it allows us to isolate the role of social preferences in a highly controlled setting, which eliminates complications that are related to decision-making under uncertainty. For instance, it is well-known that, even in one-person decision tasks, people often do not maximize expected utility (see Camerer, 1995), and that moreover, in social contexts, social and risk preferences may interact in non-trivial ways (see, e.g., Bolton and Ockenfels, 2010, and the references therein). The complete information environments in our study avoid such complicating factors.[2]

The robust mechanism which maximizes the seller's expected profits under selfish preferences has the following properties: (i) The trading surplus is allocated in an asym-

[1]Throughout we focus on social choice functions, as opposed to social choice correspondences. Consequently, by Corollary 1 in Bergemann and Morris (2005), *ex post* implementability is both necessary and sufficient for robust implementability. Moreover, if agents are selfish, then our environment gives rise to private values so that incentive compatibility in an *ex post* sense is equivalent to the requirement that truth-telling is a dominant strategy under a direct mechanism for the given social choice function.

[2]Thus, for our experimental testing strategy, we take for granted the equivalence between implementability in all complete information environments and implementability in all incomplete information environments. We explicitly investigate the former and draw conclusions for the latter. We also take for granted the validity of the revelation principle. That is, we only check whether individuals behave truthfully under a direct mechanism for a given social choice function. We discuss the advantages and limits of this approach in our concluding section.

metric way, i.e. the seller gets a larger fraction than the buyer; (ii) Whenever the buyer's valuation is low, his participation constraint binds, so that he does not realize any gains from trade; (iii) Whenever the buyer's valuation is high, his incentive constraint binds, so that he is indifferent between revealing his valuation and understating it. Experimentally, we find that under this mechanism, a non-negligible fraction of high valuation buyers understates their valuation. In all other situations, deviations — if they occur at all — are significantly less frequent.

We argue that this pattern is consistent with models of social preferences such as Fehr and Schmidt (1999), and Falk and Fischbacher (2006), among others. The basic idea is the following. A buyer with a high valuation can understate his valuation at a very small personal cost since the relevant incentive constraint binds. The benefit of this strategy is that this reduces the seller's payoff and therefore brings the seller's payoff closer to his own, thereby reducing inequality. In fact, as we will demonstrate later, many social preference models would predict this behavior.

We then introduce a class of direct mechanisms that "work" if the possibility of social preferences is acknowledged. Specifically, we introduce the notion of a direct mechanism that is externality-free. Under such a mechanism, the buyer's equilibrium payoff does not depend on the seller's type and vice versa; i.e. if, say, the buyer reveals his valuation, his payoff no longer depends on whether the seller communicates a high or a low cost to the mechanism designer. Hence, the seller cannot influence the buyer's payoff.

Almost all widely-used models of social preferences satisfy a property of selfishness in the absence of externalities, i.e. if a player considers a choice between two actions a and b, and moreover, if the monetary payoffs of everybody else are unaffected by this choice, then the player will choose a over b if her own payoff under a is higher than her own payoff under b. Now, suppose that a direct mechanism is *ex post* incentive-compatible and externality-free. Then truth-telling will be an equilibrium for any social preference model in which individuals are selfish in the absence of externalities.

We impose externality-freeness as an additional constraint on our problem of robust mechanism design, i.e. we have to design the mechanism so that it has the following property: Suppose that the traded quantity goes up because we move from a state of the world in which the seller's cost is high to a state in which the seller's cost is low. Then, there has to be an accompanying change in the price the buyer has to pay. This change needs to be calibrated in such a way that the buyer's trading surplus remains unaffected. We then characterize the optimal robust and externality-free mechanism and investigate its performance in an experiment. We find that there are no longer deviations from truth-telling. We interpret this finding as providing evidence for the relevance of social preferences in mechanism design: If there are externalities a significant fraction of individuals deviates from truth-telling. If those externalities are shut down, individuals behave truthfully.

Externality-freeness is an additional constraint. While it makes sure that individuals behave in a predictable way it reduces expected profits relative to the theoretical benchmark of a model with selfish preferences. This raises the question whether the seller makes more money if she uses an externality-free mechanism. We answer this both theoretically and empirically: The externality-free mechanism makes more money if the number of individuals whose behavior is motivated by social preferences exceeds a threshold. In our laboratory context, this number was below the threshold, so that the "conventional" mechanism made more money than the externality-free mechanism.

Based on these observations, we finally engineer a mechanism that satisfies the property of externality-freeness only locally. Specifically, we impose externality-freeness for those action-profiles where deviations from selfish behavior were frequently observed in our experiment data. We show theoretically that local externality-freeness is a constraint that can be satisfied without having to sacrifice performance: If all agents are selfish then there is an optimal mechanism that is locally externality-free. In our experiment data, however, an optimal mechanism that is locally externality-free performs significantly better than an optimal mechanism that is not externality-free. Hence, if one knows precisely which deviations from selfish behavior are tempting, one can design a mechanism that performs strictly better than both the optimal mechanism for selfish agents and the optimal globally externality-free mechanism.

Redistributive Income Taxation. The bilateral trade setup is one in which externalities are at the center of the allocation problem. Therefore, the requirement of externality -freeness may appear demanding. In settings different from the bilateral trade problem, externality-freeness may arise naturally. E.g., price-taking behavior in markets with a large number of participants gives rise to externality-freeness. If a single individual changes her demand, this leaves prices unaffected and so remain the options available to all other agents.[3] Another setting in which externality-freeness may appear natural is the design of tax systems. Here, ex- ternality-freeness requires that income taxes paid by one individual depend only on this individual's income, and not on the income earned by other individuals. Thus, when formalizing the modern approach to optimal income taxation, Mirrlees (1971) and his followers have looked exclusively at externality-free allocations.

However, as has been shown by Piketty (1993), for an economy with finitely many individuals and a commonly known cross-section distribution of types, an optimal Mirrleesian income tax system can be outperformed by one that is *not* externality-free. Specifically, Piketty shows that first-best utilitarian redistribution from high-skilled individuals to low-skilled individuals can be reached, while this is impossible with a Mirrleesian approach. A crucial feature of Piketty's approach is that types are assumed to be correlated in a particular way. For instance, if there are two individuals and it is commonly known that one of them is high-skilled and one is low-skilled, then the individu-

[3] Market behavior is therefore unaffected by social preferences, see Dufwenberg et al. (2011).

als' types are perfectly negatively correlated: If person 1 is of high ability, then person 2 is of low ability and vice versa. Piketty's construction of a mechanism that reaches the first-best utilitarian outcome heavily exploits this feature of the environment.[4]

Piketty's analysis resembles the possibility results by Crèmer and McLean (1985, 1988) in auction theory. Crèmer and McLean have shown that, with correlated values and selfish agents, there exist Bayes-Nash equilibria that achieve first-best outcomes. These findings have then be generalized to other types of allocation problems, see e.g., Kosenok and Severinov (2008). Importantly, the mechanisms that achieve first-best outcomes in the presence of correlated types give rise to payoff interdependencies or externalities among the players. Therefore, they raise the question whether social preferences might interfere with the possibility to achieve first-best outcomes. Piketty's treatment of the income tax problem is an example that allows us to get at this more general question in a particular context.

We run an experiment and show that Piketty's mechanism indeed provokes deviations from the intended behavior, and again, we argue that these deviations can be explained by models of social preferences. We then compare Piketty's mechanism to an optimal Mirrleesian mechanism. The latter is externality-free and we find that it successfully controls behavior; there are no longer significant deviations from truth-telling. We also find that the level of welfare that is generated by the Mirrleesian mechanism is significantly larger than the level of welfare generated by Piketty's mechanism. This last observation makes an interesting difference to our findings for the bilateral trade problem. Remember that, for the bilateral problem, imposing externality-freeness helped to control behavior, while the optimal mechanism for selfish agents still outperformed the optimal externality-free mechanism because, in our experiment data, the deviations from selfish behavior were not frequent enough. With the income tax problem, by contrast, imposing externality-freeness is also good for the performance of the mechanism.

Outlook. The next section discusses related literature. Section 4.3 describes the economic environment. Section 4.4 contains a detailed description of the bilateral trade problem that we study. In addition, we elaborate on why models of social preferences are consistent with the observation that individuals deviate from truth-telling under a mechanism that would be optimal if all individuals were selfish, and with the observation that they do not deviate under a mechanism that is externality-free. Section 4.5 describes our laboratory findings for the bilateral trade problem, and in Section 4.6, we clarify the conditions under which an optimal externality-free mechanism outperforms an optimal mechanism for selfish agents and relate them to our experiment data. Section 4.7 looks at an engineering approach that does impose externality-freeness only locally. Section 4.8 contains our analysis of the income tax problem. The last section concludes.

[4]If individual types are the realizations of independent random variables, then the optimal mechanism is externality-free, see Bierbrauer, 2011b.

4.2 Related literature

Our work is related to different strands of the literature. The main part of our analysis builds on the model of Myerson and Satterthwaite (1983). They establish an impossibility result for efficient trade in a setting with two privately informed parties.[5] We embed this problem into a model of robust mechanism design, see Bergemann and Morris (2005). Other contributions to the literature on robust mechanism design include Ledyard (1978), Gershkov et al. (2013) and Börgers (2015).

We also consider a problem of income taxation. The classical reference is Mirrlees (1971). We relate the Mirrleesian treatment to an alternative one that has been proposed by Piketty (1993). The mechanism design approach to the problem of optimal income taxation is also discussed in Hammond (1979), Stiglitz (1982), Dierker and Haller (1990), Guesnerie (1995), and Bierbrauer (2011b).

There is a large experimental economics literature testing mechanisms. Most laboratory studies deal with mechanisms to overcome free-riding in public goods environments (Chen, 2008), auction design (e.g., Ariely, Ockenfels, and Roth, 2005, Kagel, Lien, and Milgrom, 2010), and the effectiveness of various matching markets (e.g., Kagel and Roth, 2000, Chen and Sönmez, 2006). Roth (2012) provides a survey. Some studies take into account social preferences when engineering mechanisms. For instance, it has been shown that feedback about others' behavior or outcomes, which would be irrelevant if agents were selfish, can strongly affect social comparison processes and reciprocal interaction, and thus the effectiveness of mechanisms to promote efficiency and resolve conflicts (e.g., Chen et al., 2010, Bolton, Greiner, and Ockenfels, 2013, Ockenfels, Sliwka, and Werner, 2014; Bolton and Ockenfels, 2012 provide a survey). Social preferences are also important in bilateral bargaining with complete information, most notably in ultimatum bargaining (Güth, Schmittberger, and Schwarze, 1982; Güth and Kocher, 2014 provide a survey). In fact, this literature has been a starting point for various social preference models that we are considering in this chapter — yet the observed patterns of behavior have generally not been related to the mechanism design literature. This is different with laboratory studies of bilateral trade with incomplete information, such as Radner and Schotter (1989), Valley et al. (2002) and Kittsteiner, Ockenfels, and Trhal (2012). One major finding in this literature is, for instance, that cheap talk communication among bargainers can significantly improve efficiency. These findings are generally not related to social preference models, though.

There is a large literature on mechanism design with interdependent valuations, see e.g. the survey in Jehiel and Moldovanu (2006). In principle, models of outcomes-based social preferences such as Fehr and Schmidt (1999) or Bolton and Ockenfels (2000) can be viewed as specific models with interdependent valuations.[6] By contrast, models with

[5]Related impossibility results hold for problems of public-goods provision, see Güth and Hellwig, 1986 and Mailath and Postlewaite, 1990.

[6]When it comes to frequently used functional forms there is a difference though. The literature in mech-

intentions-based social preferences such as Rabin (1993) or Falk and Fischbacher (2006) cannot be viewed as models with interdependent valuations. In these models, preferences are menu-dependent, see Sobel (2005) for a discussion. Such a menu dependence does not arise in the literature on mechanism design with interdependent valuations. Thus, there are both similarities and important differences between these literatures. Our approach ensures that the social choice functions can be implemented irrespectively of the nature and the intensity of the individual's social preferences. It is not meant to take account of all the allocative and informational externalities that have been the focus of the literature on mechanism design with interdependent valuations.

Our study contributes to the literature by linking prominent models of social preferences with the mechanism design literature mentioned above. Bierbrauer and Netzer (2012) explore the implications of a specific model of social preferences, namely the one by Rabin (1993), for a Bayesian mechanism design problem — as opposed to a problem of robust mechanism design. They show that, to any mechanism that is incentive compatible, one can construct an "essentially" equivalent version which is externality-free and therefore should generate the intended behavior even if individuals have social preferences. Bartling and Netzer (2014) use this observation to construct an externality-free version of the second-price auction. They show experimentally that there is significant overbidding in a standard second-price auction. Overbidding disappears with the externality-free version. The work of Bartling and Netzer is related to this chapter in that it also makes use of externality-freeness. There is, however, an important difference. Since we work with ex post — as opposed to Bayesian — incentive and participation constraints, the equivalence result in Bierbrauer and Netzer (2012) no longer holds, i.e. externality-freeness becomes a substantive constraint. A contribution of this chapter is the characterization of a mechanism that is optimal in the set of those which are externality-free and ex post incentive-compatible.

4.3 The bilateral trade problem

There are two agents, referred to as the buyer and the seller. An economic outcome is a triple (q, p_s, p_b), where $q \in \mathbb{R}_+$ is the quantity that is traded, $p_b \in \mathbb{R}$ is a payment made by the buyer, and $p_s \in \mathbb{R}$ is a payment received by the seller. Monetary payoffs are $\pi_b = \theta_b q - p_b$, for the buyer and $\pi_s = -\theta_s k(q) + p_s$, for the seller where k is an increasing and convex cost function. The buyer's valuation θ_b either takes a high or a low value, $\theta_b \in \Theta_b = \{\underline{\theta}_b, \bar{\theta}_b\}$. Similarly, the seller's cost parameter θ_s can take a high or a low value so that $\theta_s \in \Theta_s = \{\underline{\theta}_s, \bar{\theta}_s\}$. A pair $(\theta_b, \theta_s) \in \Theta_b \times \Theta_s$ is referred to as a state of

anism design often focuses on preferences that are quasi-linear in money and models interdependence as arising solely through the valuations of the allocated objects. Consequently, the individuals' monetary payments do not give rise to an interdependence. By contrast, the literature on outcomes-based social preferences focuses on inequity concerns that take account of the individuals' monetary payments.

the economy. A social choice function or direct mechanism $f : \Theta_b \times \Theta_s \to \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}$ specifies an economic outcome for each state of the economy. Occasionally, we write $f = (q^f, p_b^f, p_s^f)$ to distinguish the different components of f.[7]

We denote by

$$\pi_b(\theta_b, f(\theta_b', \theta_s')) := \theta_b q^f(\theta_b', \theta_s') - p_b^f(\theta_b', \theta_s')$$

the payoff that is realized by a buyer with type θ_b if he announces a type θ_b' and the seller announces a type θ_s' under direct mechanism f. The expression $\pi_s(\theta_s, f(\theta_b', \theta_s'))$ is defined analogously.

We assume that the buyer has private information on whether his valuation θ_b is high or low. Analogously, the seller privately observes whether θ_s takes a high or a low value. Hence, a direct mechanism induces a game of incomplete information. Our analysis in the following focuses on a very specific and artificial class of incomplete information environments, namely the ones in which the types are commonly known among the players but unknown to the mechanism designer. In total there are four such complete information environments, one for each state of the economy.[8] It has been shown by Bergemann and Morris (2005) that the implementability of a social choice function in all such complete information environments is not only necessary but also sufficient for the robust implementability of a social choice function, i.e. for its implementability in all conceivable incomplete information environments. Thus, our focus on complete information environments is not only useful to cleanly isolate the effect of social preferences, but also justified by the robustness criterion.

Suppose that individuals are only interested in their own payoff. Then truth-telling is an equilibrium in all complete information environments if and only if the following ex post incentive compatibility constraints are satisfied: For all $(\theta_b, \theta_s) \in \Theta_b \times \Theta_s$,

$$\pi_b(\theta_b, f(\theta_b, \theta_s)) \geq \pi_b(\theta_b, f(\theta_b', \theta_s)) \quad \text{for all} \quad \theta_b' \in \Theta_b \ , \tag{4.1}$$

and

$$\pi_s(\theta_s, f(\theta_b, \theta_s)) \geq \pi_s(\theta_s, f(\theta_b, \theta_s')) \quad \text{for all} \quad \theta_s' \in \Theta_s \ . \tag{4.2}$$

Moreover, individuals prefer to play the mechanism over a status quo outcome with no trade if and only if the following ex post participation constraints are satisfied: For all $(\theta_b, \theta_s) \in \Theta_b \times \Theta_s$,

$$\pi_b(\theta_b, f(\theta_b, \theta_s)) \geq \bar{\pi}_b \quad \text{and} \quad \pi_s(\theta_s, f(\theta_b, \theta_s)) \geq \bar{\pi}_s \ , \tag{4.3}$$

[7] Our setting differs from the one originally studied by Myerson and Satterthwaite (1983) in that we have a convex cost function for the seller and allow for quantities in $\mathbb{R}$. In the original paper, the seller's cost function is linear and quantities are in $[0, 1]$.

[8] "Complete information" refers to a situation in which the players' monetary payoffs are commonly known. Information may still be incomplete in other dimensions, e.g., regarding the weight of fairness considerations in the other player's utility function.

where $\bar{\pi}_b$ and $\bar{\pi}_s$ are, respectively, the buyer's and the seller's payoffs in the absence of trade.

Throughout, we limit attention to direct mechanisms and to truth-telling equilibria. For models with selfish individuals, or more generally, for models with outcome-based preferences – which possibly include a concern for an equitable distribution of payoffs – this is without loss of generality by the revelation principle. For models with intention-based social preferences, such as Rabin (1993) or Dufwenberg and Kirchsteiger (2004), the revelation principle does not generally hold, see Bierbrauer and Netzer (2012) for a proof. Still, it is a sufficient condition for the implementability of a social choice function that it can be implemented as the truth-telling equilibrium of a direct mechanism. We focus on this sufficient condition, and note that it is also necessary if preferences are outcome-based.

Another property of interest to us is the externality-freeness of a social choice function f. This property holds if, for all $\theta_b \in \Theta_b$,

$$\pi_b(\theta_b, f(\theta_b, \underline{\theta}_s)) = \pi_b(\theta_b, f(\theta_b, \bar{\theta}_s)),$$

and if, for all $\theta_s \in \Theta_s$,

$$\pi_s(\theta_s, f(\underline{\theta}_b, \theta_s)) = \pi_s(\theta_s, f(\bar{\theta}_b, \theta_s)).$$

If these properties hold, then the buyer, say, cannot influence the seller's payoff, provided that the latter tells the truth. I.e. the buyer's report does not come with an externality on the seller. As we will argue later in more detail, many models of social preferences give rise to the prediction that externality-freeness in conjunction with ex post incentive compatibility is a sufficient condition for the implementability of a social choice function.

4.4 Mechanism design with and without social preferences

This section contains theoretical results which relate mechanism design theory to models of social preferences. Throughout, we use the bilateral trade problem to illustrate the conceptual questions that arise. We begin with the benchmark of optimal mechanism design under the assumption that individuals are purely selfish. We then show that many models of social preferences give rise to the prediction that such mechanisms will not generate truthful behavior. However, while maximizing expected payoffs is a well-defined goal, there are many ways to be socially motivated. In fact, one of the most robust insights from behavioral economics and psychology is the large variance of social behaviors across individuals (e.g., Camerer, 2003). As a result, there is now a plethora of social preference models, and almost all models permit individual heterogeneity by allowing different parameter values for different individuals (e.g., Cooper and Kagel, 2013). This poses a problem for mechanism design, because optimal mechanisms depend on the nature of the agents' preferences. Our approach to deal with this problem is neither to just select one of those models, nor are we even attempting to identify the best model. We will also not assume that idiosyncratic social preferences are commonly known. All these approaches would violate the spirit of robust mechanism design and the Wilson doctrine. Rather, we restrict our attention to a property of social preferences which is shared by almost all widely-used social preference models and which is independent of the exact parameter values: individuals are selfish if there is no possibility to affect the payoffs of others. As we will show, this general property of social behavior is already sufficient to construct "externality-free" mechanisms which generate truthful behavior, regardless of what is known about the specific type and parameters of the agents' social preferences.

Our approach may come at a cost. While we will be able to better control behavior than when we assume selfish preferences, not knowing the exact details of preferences may impair the profitability of the mechanism. As we show in Section 4.6, for the bilateral trade problem, the optimal robust and externality-free mechanism outperforms the optimal robust mechanisms only if the probability of behavior that is motivated by social concerns is sufficiently high.

4.4.1 Optimal mechanism design under selfish preferences

A mechanism designer wishes to come up with a mechanism for bilateral trade. Design takes place at the *ex ante* stage, i.e. before the state of the economy is realized. The designer acts in the interest of one of the parties, here the seller. The designer does not know what information the buyer and the seller have about each other at the moment where trade takes place. Hence, he seeks robustness with respect to the information structure and employs *ex post* incentive and participation constraints. The designer assumes that individuals are selfish so that these constraints are sufficient to ensure that individuals are willing to play the corresponding direct mechanism and to reveal their types. Finally, he requires budget balance only in an average sense. (Possibly, the mechanism is executed frequently, so that the designer expects to break even if budget balance holds on average.) The flexibility provided by the requirement of expected budget balance is important for some of the results that follow. With a requirement of *ex post* budget balance there would be less scope for adjusting the traded quantities and the corresponding payments to the privately held information of the buyer and the seller.[9]

[9] We do not wish to argue that the requirement of expected budget balance is, for practical purposes, more relevant than the requirement of *ex post* budget balance. This will depend on the application. The mechanisms that we study in this paper are primarily meant as diagnostic tools for the relevance of social preferences in mechanism design. In this respect, the requirement of expected budget balance proved useful.

Formally, we assume that a social choice function f is chosen with the objective to maximize expected seller profits, $\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s)\pi_s(\theta, f(\theta_b, \theta_s))$, where g is a probability mass function that gives the mechanism designer's subjective beliefs on the likelihood of the different states of the economy. The incentive and participation constraints in (4.1), (4.2) and (4.3) have to be respected. In addition, the following resource constraint has to hold

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_b^f(\theta_b, \theta_s) \geq \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_s^f(\theta_b, \theta_s) \; . \tag{4.4}$$

To solve this *full problem*, we first study a *relaxed problem* which leaves out the incentive and participation constraints for the seller. Proposition 1 characterizes its solution. This solution to the relaxed problem is also a solution to the full problem if it satisfies all constraints of the full problem. For reasons of brevity, we do not present another Proposition that clarifies the conditions under which this is the case. Instead, Example 4.1 below provides an illustration that is based on particular functional forms and parameter values.

Proposition 4.1. *A social choice function f solves the relaxed problem of robust mechanism design if and only if it has the following properties:*

(a) For any one $\theta_s \in \Theta_s$, the participation constraint of a low type buyer is binding:

$$\pi_b(\underline{\theta}_b, f(\underline{\theta}_b, \theta_s)) = \bar{\pi}_b \; .$$

(b) For any one $\theta_s \in \Theta_s$, the incentive constraint of a high type buyer is binding:

$$\pi_b(\overline{\theta}_b, f(\overline{\theta}_b, \theta_s)) = \pi_b(\overline{\theta}_b, f(\underline{\theta}_b, \theta_s)) \; .$$

(c) The trading rule is such that, for any one $\theta_s \in \Theta_s$, there is a downward distortion at the bottom

$$q^f(\underline{\theta}_b, \theta_s) \quad \in \quad \mathit{argmax}_q \left(\underline{\theta}_b - \frac{g(\overline{\theta}_b, \theta_s)}{g(\underline{\theta}_b, \theta_s)} (\overline{\theta}_b - \underline{\theta}_b) \right) q - \theta_s k(q) \; ,$$

and no distortion at the top

$$q^f(\overline{\theta}_b, \theta_s) \quad \in \quad \mathit{argmax}_q \; \overline{\theta}_b q - \theta_s k(q) \; .$$

(d) The payment rule for the buyer is such that, for any one θ_s,

$$p_b^f(\underline{\theta}_b, \theta_s) = \underline{\theta}_b q^f(\underline{\theta}_b, \theta_s) \; ,$$

and

$$p_b^f(\overline{\theta}_b, \theta_s) = \overline{\theta}_b q^f(\overline{\theta}_b, \theta_s) - (\overline{\theta}_b - \underline{\theta}_b) q^f(\underline{\theta}_b, \theta_s) \; .$$

(e) The revenue for the seller is such that

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_b^f(\theta_b, \theta_s) = \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_s^f(\theta_b, \theta_s) \ .$$

A formal proof of Proposition 4.1 is in part 4.B of the Appendix. Here, we provide a sketch of the main argument: Since we leave out the seller's incentive constraint, we can treat the seller's cost parameter as a known quantity. Hence, we think of the relaxed problem as consisting of two separate profit-maximization problems, one for a high-cost seller and one for a low-cost seller, which are linked only via the resource constraint. In each of these problems, however, the buyer's incentive and participation constraints remain relevant. Therefore, we have two profit-maximization problems. The formal structure of any one of those problems is the same as the structure of a non-linear pricing problem with two buyer types. This problem is well-known so that standard arguments can be used to derive properties (a)-(e) above.[10]

The solution to the relaxed problem leaves degrees of freedom for the specification of the payments to the seller. Consequently, any specification of the seller's revenues, so that the expected revenue is equal to the buyer's expected payment, is part of a solution to the relaxed problem. If there is one such specification that satisfies the seller's ex post incentive and participation constraints, then this solution to the relaxed problem is also a solution to the full problem. In the following we provide a specific example in which these payments are specified in such a way that they satisfy not only these constraints, but also give rise to ex post budget balance, i.e. in every state (θ_b, θ_s), the price paid by the buyer equals the revenue obtained by the seller,

$$p_b^f(\theta_b, \theta_s) = p_s^f(\theta_b, \theta_s) \ . \tag{4.5}$$

Example 4.1: An optimal robust social choice function. Suppose that $\underline{\theta}_b = 1$, $\overline{\theta}_b = 1.30$, $\underline{\theta}_s = 0.20$, and $\overline{\theta}_s = 0.65$. Also assume that the seller has a quadratic cost function $k(q) = \frac{1}{2}q^2$. Finally, assume that the reservation utility levels of both the buyer and the seller are given by $\bar{\pi}_b = \bar{\pi}_s = 2.68$. For these parameters, an optimal robust social choice function f looks as follows: The traded quantities are given by

$$q^f(\underline{\theta}_b, \underline{\theta}_s) = 3.50, \ q^f(\underline{\theta}_b, \bar{\theta}_s) = 1.08, \ q^f(\bar{\theta}_b, \underline{\theta}_s) = 6.50 \quad \text{and} \quad q^f(\bar{\theta}_b, \bar{\theta}_s) = 2.00 \ .$$

The buyer's payments are

$$p_b^f(\underline{\theta}_b, \underline{\theta}_s) = 3.50, \ p_b^f(\underline{\theta}_b, \bar{\theta}_s) = 1.08, \ p_b^f(\bar{\theta}_b, \underline{\theta}_s) = 7.40 \quad \text{and} \quad p_b^f(\bar{\theta}_b, \bar{\theta}_s) = 2.28 \ .$$

[10] A classical reference is Mussa and Rosen (1978), see Bolton and Dewatripont (2005) for a textbook treatment.

Finally, the seller's revenues are

$$p_s^f(\underline{\theta}_b, \underline{\theta}_s) = 3.50,\ p_s^f(\underline{\theta}_b, \bar{\theta}_s) = 1.08,\ p_s^f(\bar{\theta}_b, \underline{\theta}_s) = 7.40 \quad \text{and} \quad p_s^f(\bar{\theta}_b, \bar{\theta}_s) = 2.28\ .$$

By construction, f is ex post incentive compatible and satisfies the ex post participation

constraints. However, it is not externality-free. These properties can be verified by looking at the games which are induced by this social choice function on the various complete information environments. For instance, the following matrix represents the normal form game that is induced by f in a complete information environment so that the buyer has a low valuation and the seller has a low cost.[11]

Table 4.1: Low valuation buyer and low cost seller

The game induced by f for $(\theta_b, \theta_s) = (\underline{\theta}_b, \underline{\theta}_s)$.

(π_b^f, π_s^f)	$\underline{\theta}_s$	$\bar{\theta}_s$
$\underline{\theta}_b$	$(2.68, 5.52)$	$(2.68, 3.88)$
$\bar{\theta}_b$	$(1.56, 6.65)$	$(2.33, 5.03)$

The first entry in each cell is the buyer's payoff, the second entry in the cell is the seller's payoff. If both individuals truthfully reveal their types, the payoffs in the upper left corner are realized. Note that under truth-telling both payoffs are weakly larger than the reservation utility of 2.68 so that the relevant ex post participation constraints are satisfied. Also note that the seller does not benefit from an exaggeration of her cost, if the buyer communicates his low valuation truthfully. Likewise, the buyer does not benefit from an exaggeration of his willingness to pay, given that the seller communicates her low cost truthfully. Hence, the relevant ex post incentive constraints are satisfied. Finally, note that externality-freeness is violated: If the seller behaves truthfully, her payoff is higher if the buyer communicates a high willingness to pay.

For later reference, we also describe the normal form games that are induced in the remaining complete information environments.

[11] More precisely, this and the following normal form games are generated by an approximation f^x of f which is such that, whenever an incentive constraint is binding under f, a deviation from truth-telling has a small cost of two cents under f^x. Our laboratory experiments used f^x rather than f. Thus, under f^x it is less tempting to deviate from truth-telling and we can be more confident that the deviations from truth-telling that we observe reflect social preferences, as opposed to an arbitrary selection from a set of best responses.

Table 4.2: Low valuation buyer and high cost seller

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(2.68, 2.08)$	$(2.68, 3.56)$
$\overline{\theta}_b$	$(1.56, -5.23)$	$(2.33, 3.90)$

Table 4.3: High valuation buyer and low cost seller

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.97, 5.52)$	$(3.06, 3.88)$
$\overline{\theta}_b$	$(3.99, 6.65)$	$(3.08, 5.03)$

Table 4.4: High valuation buyer and high cost seller

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.97, 2.08)$	$(3.06, 3.56)$
$\overline{\theta}_b$	$(3.99, -5.23)$	$(3.08, 3.90)$

An inspection of Tables 4.1 trough 4.4 reveals the following properties of f: (i) Under truth-telling the seller's payoff exceeds the buyer's payoff in all states of the economy, (ii) if the buyer's type is low (Tables 4.1 and 4.2), then his payoff under truth-telling is equal to his reservation utility level of 2.68, i.e. the participation constraint of a low type buyer binds, (iii) if the buyer's type is high (Tables 4.3 and 4.4), then the buyer's incentive constraint is binding in the sense that understating comes at a very small personal cost (the payoff drops from 3.99 to of 3.97).

4.4.2 An observation on models of social preferences

We now show that the social choice function in Proposition 4.1 is not robust in the following sense: It provokes deviations from truth-telling if individuals are motivated by social preferences. To formalize a possibility of social preferences, we assume that any one individual $i \in \{b, s\}$ has a utility function $U_i(\theta_i, r_i, r_i^b, r_i^{bb})$ which depends in a parametric way on the individual's true type θ_i and, in addition, on the following three arguments: the individual's own report r_i, the individual's (first order) belief about the other player's report, r_i^b, and the individuals' (second order) belief about the other player's first-order belief, r_i^{bb}. Different models of social preferences make different assumptions about these utility functions.

Intention-based social preferences. Second-order beliefs play a role in models with intention-based social preferences such as Rabin (1993), Dufwenberg and Kirchsteiger (2004) or Falk and Fischbacher (2006). In these models, the utility function takes the following form

$$U_i(\theta_i, r_i, r_i^b, r_i^{bb}) = \pi_i(\theta_i, f(r_i, r_i^b)) + y_i\, \kappa_i(r_i, r_i^b, r_i^{bb})\, \kappa_j(r_i^b, r_i^{bb}) \; . \tag{4.6}$$

The interpretation is that the players' interaction gives rise to sensations of kindness or unkindness, as captured by $y_i\, \kappa_i(r_i, r_i^b, r_i^{bb})\, \kappa_j(r_i^b, r_i^{bb})$. In this expression, $y_i \geq 0$ is an exogenous parameter, interpreted as the weight that agent i places on kindness considerations. The term $\kappa_i(r_i, r_i^b, r_i^{bb})$ is a measure of how kindly i intends to treat the other agent j. While the models of Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) differ in some respects, they all make the following assumption: Given r_i^b and r_i^{bb}, for any two reports r_i' and r_i'', $\pi_j(\theta_j, f(r_i', r_i^b)) \geq \pi_j(\theta_j, f(r_i'', r_i^b))$ implies that $\kappa_i(r_i', r_i^b, r_i^{bb}) \geq \kappa_i(r_i'', r_i^b, r_i^{bb})$, i.e. the kindness intended by i is larger if her report yields a larger payoff for j. Second-order beliefs are relevant here if player i expresses kindness by increasing j's payoff relative to the payoff that, according to the beliefs of i, j expects to be realizing. The latter payoff depends on the beliefs of i about the beliefs of j about i's behavior.

Whether or not i's utility is increasing in κ_i depends on i's belief about the kindness that is intended by player j and which is denoted by κ_j. If $\kappa_j > 0$, then i believes that j is kind and her utility increases, ceteris paribus, if j's payoff goes up. By contrast, if $\kappa_j < 0$, then i believes that j is unkind and her utility goes up if j is made worse off. Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) all assume that the function κ_j is such that, for given second-order beliefs r_i^{bb}, $\kappa_j(r_i^{b\prime}, r_i^{bb}) \geq \kappa_i(r_i^{b\prime\prime}, r_i^{bb})$ whenever $\pi_i(\theta_i, f(r_i^{b\prime}, r_i^{bb})) \geq \pi_i(\theta_i, f(r_i^{b\prime\prime}, r_i^{bb}))$. Second-order beliefs play a role here because, in order to assess the kindness that is intended by j, i has to form a belief about j's belief about i's report.

Outcome-based social preferences. In models with outcome-based social preferences such as Fehr and Schmidt (1999), Bolton and Ockenfels (2000), or Charness and Rabin (2002) second order beliefs play no role and individuals are assumed to care about their own payoff and the distribution of payoffs among the players. For instance, with Fehr-Schmidt-preferences, the utility function of individual i reads as

$$\begin{aligned} U_i(\theta_i, r_i, r_i^b, r_i^{bb}) \;\; = \;\; & \pi_i(\theta_i, f(r_i, r_i^b)) \\ & -\alpha_i \max\{\pi_j(\theta_j, f(r_i, r_i^b)) - \pi_i(\theta_i, f(r_i, r_i^b)), 0\} \\ & -\beta_i \max\{\pi_i(\theta_i, f(r_i, r_i^b)) - \pi_j(\theta_j, f(r_i, r_i^b)), 0\} \; , \end{aligned} \tag{4.7}$$

where it is assumed that $\alpha_i \geq \beta_i$ and that $0 \leq \beta_i < 1$.

Implications for the social choice function in Proposition 4.1. Many models of social preferences give rise to the prediction that a social choice function that would be optimal if individual were selfish will trigger deviations from truth-telling. Specifically, for our bilateral trade problem, high valuation buyers will understate their valuation. Models of outcome-based and intention-based social preferences provide different explanations for this: With outcome-based social preferences, the buyer may wish to harm the seller so as to make their expected payoffs more equal. The reasoning for intention-based models, such as Rabin (1993), would have a different logic. For the game in Table 4.4, the buyer would argue as follows: My expected payoff would be higher if the seller deviated from truth-telling and communicated a low cost. Since the seller does not make use of this opportunity to increase my payoff, he is unkind. I therefore wish to reciprocally reduce his expected payoff.

Whatever the source of the desire to reduce the seller's payoff, a high valuation buyer can reduce the seller's payoff by understating his valuation. Since the relevant incentive constraint binds, such an understatement is costless for the buyer, i.e. he does not have to sacrifice own payoff if he wishes to reduce the seller's payoff.

The following observation states this more formally for the case of Fehr-Schmidt-preferences. In Appendix 4.A, we present analogous results for other models of social preferences.

Observation 4.1. *Consider a complete information types space for state (θ_b, θ_s) and suppose that $\theta_b = \bar{\theta}_b$. Suppose that f is such that*

$$\pi_s(\theta_s, f(\overline{\theta}_b, \theta_s)) > \pi_s(\theta_s, f(\underline{\theta}_b, \theta_s)) > \pi_b(\overline{\theta}_b, f(\underline{\theta}_b, \theta_s)) = \pi_b(\overline{\theta}_b, f(\overline{\theta}_b, \theta_s)) \tag{4.8}$$

Suppose that the seller behaves truthfully. Also suppose that the buyer has Fehr-Schmidt-preferences as in (4.7) with $\alpha_b \neq 0$. Then the buyer's best response is to understate his valuation.

The social choice function in Example 4.1 fulfills Condition (4.8). Consider Tables 4.3 and 4.4. The buyer's incentive constraint binds. Moreover, if the buyer understates his valuation, this harms the seller. The harm is, however, limited in the sense that the seller's reduced payoff still exceeds the buyer's payoff. For such a situation the Fehr-Schmidt model of social preferences predicts that the buyer will deviate from truth-telling, *for any* pair of parameters (α_b, β_b) so that $\alpha_b \neq 0$. Put differently, truth-telling is a best response for the buyer only if $\alpha_b = 0$, i.e. only if the buyer is selfish.

4.4.3 Social-preference-robust mechanisms

The models of social preferences mentioned so far differ in many respects. They are, however, all consistent with the following assumption of *selfishness in the absence of externalities.*

Assumption 4.1. *Given r_i^b and r_i^{bb}, if r_i' and r_i'' are such that $\pi_j(\theta_j, f(r_i', r_i^b)) = \pi_j(\theta_j, f(r_i'', r_i^b))$ and $\pi_i(\theta_i, f(r_i', r_i^b)) > \pi_i(\theta_i, f(r_i'', r_i^b))$, then $U_i(\theta_i, r_i', r_i^b, r_i^{bb}) \geq U_i(\theta_i, r_i'', r_i^b, r_i^{bb})$.*

Assumption 4.1 holds provided that individuals prefer to choose strategies that increase their own payoff, whenever they can do so without affecting others. This does not preclude a willingness to sacrifice own payoff so as to either increase or reduce the payoff of others. It is a ceteris paribus assumption: In the set of strategies that have the same implications for player j, player i weakly prefers the ones that yield a higher payoff for herself. Assumption 4.1 has the following implication: In situations where players do not have the possibility to affect the payoffs of others, social preferences will be behaviorally irrelevant, and the players act as if they were selfish payoff maximizers.

The following observation illustrates that the utility function underlying the Fehr and Schmidt (1999)-model of social preferences satisfies Assumption 4.1 for all possible parametrization of the model. Appendix 4.A. confirms this observation for other models of social preferences.[12]

Observation 4.2. *Suppose the buyer and the seller have preferences as in (4.7) with parameters (α_b, β_b) and (α_s, β_s), respectively. The utility functions U_b and U_s satisfy Assumption 4.1, for all (α_b, β_b) so that $\alpha_b \geq \beta_b$ and $0 \leq \beta_b < 1$ and for all (α_s, β_s) so that $\alpha_s \geq \beta_s$ and $0 \leq \beta_s < 1$.*

We now define a mechanism that is robust in the following sense: For any individual i, given correct first- and second-order beliefs, a truthful report maximizes U_i, for all utility functions satisfying Assumption 4.1.

Definition 4.1. *A direct mechanism for social choice function f is said to be social-preference-robust if it satisfies the following property: On any complete information environment, given correct first- and second-order beliefs, truth-telling by any player $i \in \{b, s\}$ is a best response to truth-telling by player $j \neq i$, for all utility functions U_i satisfying Assumption 4.1.*

Social-preference-robustness of a mechanism is an attractive property. It is robust against widely varying beliefs of the mechanism designer about what is the appropriate specification and intensity of social preferences across individuals. As long as preferences satisfy Assumption 4.1, we can be assured that individuals behave truthfully under such a mechanism.

[12] Assumption 4.1 is also satisfied in models of pure altruism, see Becker (1974). All parameterized versions that Bolton and Ockenfels (2000) propose for their model are consistent with Assumption 4.1, too, although we note that it is theoretically possible to construct preferences that are consistent with their general assumptions and may still violate Assumption 4.1. Such preferences would be the only possible exception that we encountered among prominent social preference models.

The following Proposition justifies our interest in externality-free mechanisms. If we add externality-freeness to the requirement of incentive compatibility, we arrive at a social-preference-robust mechanism.

Proposition 4.2. *Suppose that f is ex post incentive-compatible and externality-free. Then f is social-preference-robust.*

Proof. Consider a complete information environment for types (θ_i, θ_j). Suppose that player i believes that player j acts truthfully so that $r_i^b = \theta_j$ and that he believes that player j believes that he acts truthfully so that $r_i^{bb} = \theta_i$. By ex post incentive compatibility, $\pi_i(\theta_i, f(r_i, r_i^b))$ is maximized by choosing $r_i = \theta_i$. By externality-freeness, $\pi_j(\theta_j, f(r_i', r_i^b)) = \pi_j(\theta_j, f(r_i'', r_i^b))$ for any pair $r_i', r_i'' \in \Theta_i$. Hence, by Assumption 4.1, $r_i = \theta_i$ solves $\max_{r_i \in \Theta_i} U_i(\theta_i, r_i, r_i^b, r_i^{bb})$. □

4.4.4 Optimal robust and externality-free mechanism design

We now add the requirement of externality-freeness to our mechanism design problem. To characterize the solution of this mechanism design problem it is instructive to begin, again, with a relaxed problem in which only a subset of all constraints is taken into account. Specifically, the relevant constraints are: the resource constraint in (4.4), the participation constraints for a low valuation buyer,

$$\pi_b(\underline{\theta}_b, f(\underline{\theta}_b, \theta_s)) \geq \bar{\pi}_b, \quad \text{for all} \quad \theta_s \in \Theta_s ,$$

the incentive constraint for a high type buyer who faces a low cost seller,

$$\pi_b(\bar{\theta}_b, f(\bar{\theta}_b, \underline{\theta}_s)) \geq \pi_b(\bar{\theta}_b, f(\underline{\theta}_b, \underline{\theta}_s)) ,$$

and, finally, the externality-freeness condition for a high valuation buyer

$$\pi_b(\bar{\theta}_b, f(\bar{\theta}_b, \underline{\theta}_s)) = \pi_b(\bar{\theta}_b, f(\bar{\theta}_b, \bar{\theta}_s)) .$$

Again, we do not provide a complete characterization of the conditions under which the solution to the relaxed problem and the solution to the full problem coincide. Example 4.2 below provides an illustration of such a case.

Proposition 4.3. *A social choice function f' solves the relaxed problem of robust and externality-free mechanism design if and only if it has the following properties:*

(a)' For any one $\theta_s \in \Theta_s$, the participation constraint of a low type buyer is binding:

$$\pi_b(\underline{\theta}_b, f'(\underline{\theta}_b, \theta_s)) = \bar{\pi}_b .$$

(b)' For $\theta_s = \underline{\theta}_s$, the incentive constraint of a high type buyer is binding.

(c)' The trading rule is such that there is a downward distortion only for state $(\underline{\theta}_b, \underline{\theta}_s)$*;*

$$q^{f'}(\underline{\theta}_b, \underline{\theta}_s) \quad \in \quad argmax_q \left(\underline{\theta}_b - \frac{g^m(\overline{\theta}_b)}{g(\underline{\theta}_b, \underline{\theta}_s)} (\overline{\theta}_b - \underline{\theta}_b) \right) q - \theta_s k(q) \ ,$$

where $g^m(\overline{\theta}_b) := g(\overline{\theta}_b, \underline{\theta}_s) + g(\overline{\theta}_b, \overline{\theta}_s)$*. Otherwise, there is no distortion.*

(d)' The payment rule for the buyer is such that, for any one θ_s*,*

$$p_b^{f'}(\underline{\theta}_b, \theta_s) = \underline{\theta}_b q^{f'}(\underline{\theta}_b, \theta_s) \ .$$

In addition

$$p_b^{f'}(\overline{\theta}_b, \underline{\theta}_s) = \overline{\theta}_b q^{f'}(\overline{\theta}_b, \underline{\theta}_s) - (\overline{\theta}_b - \underline{\theta}_b) q^{f'}(\underline{\theta}_b, \underline{\theta}_s) \ ,$$

and

$$p_b^{f'}(\overline{\theta}_b, \overline{\theta}_s) = \overline{\theta}_b q^{f'}(\overline{\theta}_b, \overline{\theta}_s) - (\overline{\theta}_b - \underline{\theta}_b) q^{f'}(\underline{\theta}_b, \underline{\theta}_s) \ ,$$

(e)' The revenue for the seller is such that

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_b^{f'}(\theta_b, \theta_s) = \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_s^{f'}(\theta_b, \theta_s) \ .$$

A formal proof of the Proposition is relegated to part 4.B of the Appendix. It proceeds as follows: The first step is to show that all inequality constraints of the relaxed problem have to be binding. Otherwise, it would be possible to implement the given trading rule $q^{f'}$ with higher payments of the buyer. This establishes $(a)'$ and $(b)'$. Second, we solve explicitly for the payments of the buyer as a function of the trading rule $q^{f'}$ — this yields $(d)'$ — and substitute the resulting expressions into the objective function. This resulting unconstrained optimization problem has first order conditions which characterize the optimal trading rule, see the optimality conditions in $(c)'$.

After having obtained the solution to the relaxed problem, we need to make sure that it is also a solution to the full problem. For the buyer, it can be shown that the neglected participation, incentive and externality-freeness constraints are satisfied provided that the solution to the relaxed problem is such that the traded quantity increases in the buyer's valuation and decreases in the seller's cost. If there is a solution to the relaxed problem that satisfies the seller's incentive, participation and externality-freeness constraints, then this solution to the relaxed problem is also a solution to the full problem. The social choice function f' in Example 4.2 below has all these properties.

The substantive difference between the optimal robust mechanism in Proposition 4.1 and the optimal robust and externality-free mechanism in Proposition 4.3 is in the pattern of distortions. The optimal robust mechanism has downward distortions whenever the buyer has a low valuation. The optimal robust and externality-free mechanism has a downward distortion in only one state, namely the state in which the buyer's valuation is low and the seller's cost is low. This distortion, however, is more severe than the distortion that arises for this state with the optimal robust mechanism.

Example 4.2: An optimal robust and externality-free social choice function. Suppose the parameters of the model are as in Example 4.1. The social choice function f', specified in Proposition 4.3, solves the problem of optimal robust and externality-free mechanism design formally defined in the previous paragraph: The traded quantities are given by

$$q^{f'}(\underline{\theta}_b, \underline{\theta}_s) = 2,\ q^{f'}(\underline{\theta}_b, \bar{\theta}_s) = 1.54,\ q^{f'}(\bar{\theta}_b, \underline{\theta}_s) = 6.5 \quad \text{and} \quad q^{f'}(\bar{\theta}_b, \bar{\theta}_s) = 2\ .$$

The buyer's payments are

$$p_b^{f'}(\underline{\theta}_b, \underline{\theta}_s) = 2,\ p_b^{f'}(\underline{\theta}_b, \bar{\theta}_s) = 1.54,\ p_b^{f'}(\bar{\theta}_b, \underline{\theta}_s) = 7.85 \quad \text{and} \quad p_b^{f'}(\bar{\theta}_b, \bar{\theta}_s) = 2\ .$$

Finally, the seller's revenues are

$$p_s^{f'}(\underline{\theta}_b, \underline{\theta}_s) = 2.52,\ p_s^{f'}(\underline{\theta}_b, \bar{\theta}_s) = 1.99,\ p_s^{f'}(\bar{\theta}_b, \underline{\theta}_s) = 6.35 \quad \text{and} \quad p_s^{f'}(\bar{\theta}_b, \bar{\theta}_s) = 2.52\ .$$

To illustrate the property of externality-freeness, we consider, once more, the various complete information games which are associated with this social choice function.

Table 4.1': The game induced by f' for $(\theta_b, \theta_s) = (\underline{\theta}_b, \underline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(2.68, 5.33)$	$(2.68, 4.86)$
$\overline{\theta}_b$	$(0.97, 5.33)$	$(2.66, 5.31)$

Along the same lines as for Table 4.1, one may verify that the relevant ex post incentive and participation constraints are satisfied. In addition, externality-freeness holds: If the seller communicates her low cost truthfully, then she gets a payoff of 5.33 irrespectively of whether the buyer communicates a high or a low valuation. Also, if the buyer reveals his low valuation, he gets 2.68 irrespectively of whether the seller communicates a high or a low cost.

Again, we also describe the normal form games that are induced by f' in the remaining complete information environments.

Table 4.2': The game induced by f' for $(\theta_b, \theta_s) = (\underline{\theta}_b, \overline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(2.68, 4.19)$	$(2.68, 4.21)$
$\overline{\theta}_b$	$(0.97, -6.57)$	$(2.66, 4.21)$

Table 4.3': The game induced by f' for $(\theta_b, \theta_s) = (\overline{\theta}_b, \underline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.41, 5.33)$	$(3.24, 4.86)$
$\overline{\theta}_b$	$(3.43, 5.33)$	$(3.43, 5.31)$

Table 4.4': The game induced by f' for $(\theta_b, \theta_s) = (\overline{\theta}_b, \overline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.41, 4.19)$	$(3.24, 4.21)$
$\overline{\theta}_b$	$(3.43, -6.57)$	$(3.43, 4.21)$

On top of externality-freeness, the social choice function f' in Tables 4.1' to 4.4' has the following properties: (i) The seller's payoff under truth-telling is higher than the buyer's payoff under truth-telling, (ii) a low type buyer realizes his reservation utility (see Tables 4.1' and 4.2'), and (iii) the buyer's incentive constraint binds if the seller's cost is low, but not if the seller's cost is high (see Tables 4.3' and 4.4').

4.5 A laboratory experiment

We conducted a laboratory experiment with five treatments. The first treatment is based on the optimal mechanism f under selfish preferences in Example 4.1 (T1), and the second treatment is based on the optimal externality-free mechanism f' under social preferences in Example 4.2 (T2). The three additional treatment variations (T3-5) will be described in subsequent sections. All treatments were conducted employing exactly the same laboratory procedures which are described below.

Laboratory Procedures. The experiments were conducted in the *Cologne Laboratory for Economic Research* at the University of Cologne. They had been programmed with *z-Tree* developed by Fischbacher (2007), and participants were recruited with the online recruitment system *ORSEE* developed by Greiner (2004). In total, we recruited 632 subjects who participated in twenty sessions, four for each of the five treatments. Each subject was allowed to participate in one session and in one treatment only (between-subject design). We collected at least 63 independent observations for each treatment and player role. Subjects were students from all faculties of the University of Cologne, mostly female (380 subjects), with an average age of 24 years. A session lasted 45-60 minutes. Average payments to subjects, including the show-up fee, was 10.76 Euro.

Upon arrival, subjects were randomly assigned to computer-terminals and received identical written instructions, which informed them about all general rules and procedures of the experiment. All treatment- and role-specific information was given on the computer-screen (see Appendix 4.C for instructions and screenshot). We used neutral terms to describe the game; e.g., player-roles were labeled Participant A (B) and strategies were labeled Top (Left) and Bottom (Right) respectively. [13]

Subjects then went through a *learning stage*, with no interaction among subjects and no decision-dependent payments. In the learning stage, subjects had to choose actions for each player role in each complete information game and then to state the resulting payoffs for the corresponding self-selected strategy combination. Subjects had to give the right answer before proceeding to the decision stage. This way we assured that all subjects were able to correctly read the payoff tables, without suggesting specific actions which might create anchoring or experimenter demand effects.

Then subjects entered the *decision stage* and were informed about their role in their matching group. The matching into groups and roles was anonymous, random and held constant over the course of the experiment. Within the decision stage, subjects had to choose one action for each of the four complete information games of their specific treatment.[14] The order of the four games was identical to the order in Table 4.5. Only after all subjects submitted their choices, feedback was given to each subject on all choices and resulting outcomes in their group. Finally, one of the four games was randomly determined for being paid in addition to a 2.50 Euro show-up fee.

Results. Table 4.5 (on page 163) summarizes the decisions made in the experiment. Sellers report their true valuation in almost all cases. Buyers with a low type also make truthful reports in both treatments. This is different for high type buyers in T1, though. Here, 13% (16%) of the buyers understate their true valuation when facing a seller with

[13] In the following we refer to the specific roles within the experiment as buyers and sellers, to make this section consistent with previous ones.

[14] As mentioned before, our experimental testing strategy takes for granted the equivalence between implementability in all complete information environments and implementability in all incomplete information environments; see Section 4.9 for discussion.

a low (high) valuation.

This pattern of buyer and seller behavior is in line with models of social preferences. In particular, for T1, which is based on an optimal mechanism for selfish agents, these models imply that high type buyers cannot be expected to always make truthful reports. By contrast, for $T2$, which is based on an optimal externality-free mechanism, these models unambiguously predict truthful behavior. We observe significantly higher shares of truthful high type buyer reports in T2 in comparison to T1 (two-sided Fisher's exact test, $p = 0.017$ for the games with a low type seller and $p = 0.014$ for the games with a high type seller).

Table 4.5: Choice Data T1 and T2

		Buyer		*Seller*	
	Game induced by	$\underline{\theta}_b$	$\overline{\theta}_b$	$\underline{\theta}_s$	$\overline{\theta}_s$
T1 *optimal mechanism under selfish preferences*	f for $(\underline{\theta}_b, \underline{\theta}_s)$	63	0	63	0
	f for $(\underline{\theta}_b, \overline{\theta}_s)$	63	0	0	63
	f for $(\overline{\theta}_b, \underline{\theta}_s)$	8	55	63	0
	f for $(\overline{\theta}_b, \overline{\theta}_s)$	10	53	1	62
T2 *externality-free mechanism*	f' for $(\underline{\theta}_b, \underline{\theta}_s)$	64	0	62	2
	f' for $(\underline{\theta}_b, \overline{\theta}_s)$	64	0	0	64
	f' for $(\overline{\theta}_b, \underline{\theta}_s)$	1	63	64	0
	f' for $(\overline{\theta}_b, \overline{\theta}_s)$	2	62	0	64

4.6 Which mechanism is more profitable?

We now turn to the question which of the two mechanisms the designer would prefer. We first clarify the conditions under which the optimal mechanism for selfish agents outperforms the optimal externality-free mechanism in the sense that it yields a higher value of the designer's objective, here, maximal expected profits for the seller. We then check whether these conditions are satisfied in our experiment data.

Based on our experiment results, we introduce a distinction between different *behavioral types* of buyers. There is the "truthful type" and the "understatement type".[15] The former communicates his valuation truthfully in all the complete information games induced by the optimal robust mechanism f. The latter communicates a low valuation in

[15] We refer to behavioral types because we wish to remain agnostic with respect to the social preference model that generates this behavior. Truthful behavior, for instance, can be rationalized both by selfish preferences and by preferences that include a concern for welfare. In the latter case, understatement is not attractive because it is Pareto-damaging.

all such games. We assume throughout that the seller always behaves truthfully, which is also what we observed in the experiment. We denote the probability that a buyer is of the "truthful type" by σ. We denote by $\Pi^f(\sigma)$ the expected profits that are realized under f. We denote by $\Pi^{f'}$ the expected profits that are realized under the optimal externality-free social choice function f', under the assumption that the buyer and the seller behave truthfully in all complete information games.

Proposition 4.4. *Suppose that* $\Pi^f(0) < \Pi^{f'}$. *Then there is a critical value* $\hat{\sigma}$ *so that* $\Pi^f(\sigma) \geq \Pi^{f'}$ *if and only if* $\sigma \geq \hat{\sigma}$.

Proof. We first note that

$$\begin{aligned} \Pi^f(\sigma) &= \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) \Big\{ \sigma(p_b^f(\theta_b, \theta_s) - \theta_s k(q^f(\theta_b, \theta_s))) \\ &\qquad\qquad + (1-\sigma)(p_b^f(\underline{\theta}_b, \theta_s) - \theta_s k(q^f(\underline{\theta}_b, \theta_s))) \Big\} \\ &= \sigma \Pi^f(1) + (1-\sigma)\Pi^f(0) \ . \end{aligned}$$

We also note that $\Pi^f(1) > \Pi^{f'}$ since $\Pi^f(1)$ gives expected profits if there are only truthful buyer types, which is the situation in which f is the optimal mechanism. The term $\sigma\Pi^f(1) + (1-\sigma)\Pi^f(0)$ is a continuous function of σ. It exceeds $\Pi^{f'}$ for σ close to one. If $\Pi^f(0) < \Pi^{f'}$, it falls short of $\Pi^{f'}$ for σ close to zero. Hence, there is $\hat{\sigma} \in (0, 1)$ so that $\Pi^f(\sigma) = \sigma\Pi^f(1) + (1-\sigma)\Pi^f(0)$ exceeds $\Pi^{f'}$ if and only if σ exceeds $\hat{\sigma}$. □

Our experiment data revisited. For the Examples 4.1 and 4.2 on which our experiments were based, the premise of Proposition 4.4 that $\Pi^f(0) < \Pi^{f'}$ is fulfilled. Specifically,

$$\Pi^f(0) = 4.54 \ , \ \Pi^f(1) = 4.91 \ , \ \Pi^f(\sigma) = 4.91 - 0.37\sigma \ , \ \Pi^{f'} = 4.77 \ \text{ and } \hat{\sigma} = 0.622$$

Thus, the fraction of deviating buyers must rise above 0.38 if the optimal external- ity-free mechanism is to outperform the optimal robust mechanism. In our experiment data, however, the fraction of deviating buyer types was only 0.14. As a consequence, actual average seller profits are smaller under the externality-free mechanism (4.77) than under the optimal robust mechanism (4.82). This difference was not found to be statistically significant, though (two-sided t-test based on independent average profits, $p = 0.143$).[16]

One might have expected more deviations from truth-telling. For instance, the social preference model by Fehr and Schmidt (1999) is consistent with truthful buyers only for

[16] Intuitively, the sellers matched with the low valuation buyers dampen the effect of the deviant high valuation buyers on the performance measure.

one special case, namely the case in which buyers are completely selfish so that $\alpha_b = 0$, and Fehr and Schmidt estimate that often roughly 50% of subjects behave in a fair way. This would have been more than enough to make the externality-free mechanism more profitable. However, the degree of selfishness may vary with the framing of the context, size of payments, etc., and moreover not all social preference models predict deviations. For instance, according to the model of Charness and Rabin (2002), individuals have a concern for welfare, so that an efficiency-damaging action such as communicating a low valuation instead of high valuation seems less attractive. This uncontrolled uncertainty about the mix of preferences among negotiators in a specific context justifies our approach to not further specify (beliefs about) social preferences.

That said, an important insight is that the ability to control behavior is not the same as the ability to reach a given objective, here maximal seller profits. Under an externality-free mechanism deviations from truth-telling are no longer tempting, i.e. this mechanism successfully controls behavior. One may, however, still prefer to use a mechanism under which some agents deviate if the complimentary set of agents who do not deviate is sufficiently large.

4.7 Finding a superior mechanism: An engineering approach

Our laboratory findings suggest that the requirement of externality-freeness is more than what is really needed to control behavior. Under the optimal mechanism for selfish agents only particular deviations from truth-telling were observed frequently: Some of the high valuation buyers understated their valuation. However, when we impose externality-freeness we also ensure that low valuation buyers do not overstate their valuation, that high cost sellers do not understate their costs, and that low cost sellers do not exaggerate their costs. While such deviations could possibly be rationalized by models of social preferences, they seem empirically less likely.

In the following, we therefore consider a mechanism design problem in which the requirement of externality-freenes is imposed only locally, namely such that the buyer is unable to influence the seller's payoff. Formally, we require that, for all $\theta_s \in \Theta_s$,

$$\pi_s(\theta_s, f(\underline{\theta}_b, \theta_s)) = \pi_s(\theta_s, f(\bar{\theta}_b, \theta_s)) \; . \tag{4.9}$$

We do not attempt to provide an axiomatic foundation for these constraints. The motivation for imposing them comes exclusively from the behavior that we observed in our laboratory tests of the mechanisms in Examples 4.1 and 4.2. This is why refer to this approach as "engineering" (see Roth, 2002 and Bolton and Ockenfels, 2012).

Remember that the optimal social choice function that we characterize in Proposition 4.1 leaves degrees of freedom for the specification of the payments to the seller.

For instance, these payments can be chosen so that, in each state, the payment of the buyer equals the seller's revenue, as stipulated by (4.5). Alternatively, the payments can be chosen so that the local externality-freeness condition (4.9) is satisfied; i.e. there is an optimal mechanism that satisfies (4.9). Hence, local externality-freeness can be ensured without having to sacrifice performance. This is stated formally in the following Proposition that we prove in part 4.B of the Appendix.

Proposition 4.5. *There is a solution to the relaxed problem of robust mechanism design, characterized in Proposition 4.1, that satisfies (4.9).*

Proposition 4.5 shows that if everybody is selfish then an optimal mechanism that satisfies ex post budget balance (as in Example 4.1 above) and an optimal mechanism that satisfies local externality-freeness (as in Example 4.3 below) are equivalent in terms of the expected profits that they generate. However, if the locally externality-free mechanism eliminates deviations that occur under ex post budget balance, it will perform strictly better. To see whether this is indeed the case we ran another laboratory treatment (T3), employing the same procedures as outlined in Section 4.5, yet based on the following Example.

Example 4.3: An optimal robust and locally externality-free social choice function. We illustrate Proposition 4.5 in the context of our numerical example. The payoff functions, parameter values, and traded quantities are as in Example 4.1. We denote the optimal mechanism that is locally externality-free by f''. Under f'', payments to the seller are given by

$$p_s^{f''}(\underline{\theta}_b, \underline{\theta}_s) = 4.065,\ p_s^{f''}(\underline{\theta}_b, \bar{\theta}_s) = 1.250,\ p_s^{f''}(\bar{\theta}_b, \underline{\theta}_s) = 6.985\ , p_s^{f''}(\bar{\theta}_b, \bar{\theta}_s) = 2.110\ .$$

Part 4.D of the Appendix contains a detailed description of the normal form games that f'' induces on the four different complete information type spaces. It also contains a detailed description of the experiment results. They can be summarized as follows: As predicted, all low valuation buyers communicated their types truthfully, just as in T1. For the states with high valuation buyers the locally externality-free mechanism has less deviations from truth-telling than the mechanism in Example 4.1. The difference is significantly different from zero for the states with a low type seller (two-sided Fisher's exact test, $p = 0.033$). It therefore also generates higher expected seller profits ($\Pi^{f''} = 4.90$) than both the mechanism in Example 4.1 ($\Pi^f = 4.82$) and the globally externality-free mechanism in Example 4.2 ($\Pi^{f'} = 4.77$). Both welfare comparisons are statistically significant (two-sided t-test, $p_{T1\,vs.\,T3} = 0.037$ and $p_{T2\,vs.\,T3} < 0.001$).

4.8 Redistributive income taxation

As in our analysis of the bilateral trade problem, we consider an economy with two individuals, $I = \{1, 2\}$. Individual i derives utility from private goods consumption, or after-tax-income, c_i, and dislikes productive effort. Individual i's productive effort is measured by $\frac{y_i}{\omega_i}$, where y_i denotes the individual's contribution to the economy's output, or pre-tax-income, and ω_i is a measure of the individual's productive abilities. Thus, an individual with high productive abilities can generate a given level of output with less effort than an individual with low productive abilities. We assume that individual preferences can be represented by an additively separable utility function

$$u(c_i) - v\left(\frac{y_i}{\omega_i}\right) ,$$

where u is an increasing and concave function, and v is an increasing and convex function. Both functions are assumed to satisfy the usual Inada conditions. Note that the individuals' preferences satisfy the single-crossing property: For any point in a (y, c)-diagram the indifference curve of an individual with low abilities through this point is steeper than the indifference curve of an individual with high abilities.

We assume that ω_i is the realization of a random variable that is privately observed by individual i. This random variable either takes a high value, ω_h, or a low value, ω_l. A state of the economy is a pair $\omega = (\omega_1, \omega_2)$ which specifies the productive ability of individual 1 and the productive ability of individual 2. The set of states is equal to $\{\omega_l, \omega_h\}^2$. A social choice function or direct mechanism consists of functions $c_i : \{\omega_l, \omega_h\}^2 \to \mathbb{R}_+$ and $y_i : \{\omega_l, \omega_h\}^2 \to \mathbb{R}_+$ which specify, for each state of the economy, and for each individual, a consumption and an output level.

An important benchmark is the first-best utilitarian welfare optimum. This is the social choice function which is obtained by choosing, separately for each state ω, $c_1(\omega)$, $c_2(\omega)$, $y_1(\omega)$ and $y_2(\omega)$ so as to maximize the sum of utilities,

$$u(c_1(\omega)) - v\left(\frac{y_1(\omega)}{\omega_1}\right) + u(c_2(\omega)) - v\left(\frac{y_2(\omega)}{\omega_2}\right) ,$$

subject to the economy's resource constraint,

$$c_1(\omega) + c_2(\omega) \leq y_1(\omega) + y_2(\omega) .$$

For a state where one individual is high-skilled and one is low-skilled this has the following implication: Both individuals get the same consumption level because marginal consumption utilities ought to be equalized. However, the high-skilled individual has to deliver more output than the low-skilled individual because marginal costs of effort ought to be equalized as well.

It will prove useful to have specific notation which refers to the first-best utilitarian welfare maximum for an economy with one highly productive and one less productive individual. The former is assigned an income requirement of y_h^* and a consumption level of c_h^*. The latter gets a lower income requirement, denoted by y_l^*, but receives the same consumption level $c_l^* = c_h^*$.

This social choice function raises questions of incentive compatibility. Clearly, the high-skilled individual would prefer the outcome intended for the low-skilled individual since the latter has the same consumption level but a smaller workload. As we will describe in the following, whether or not the first-best utilitarian welfare optimum can be reached in the presence of private information on productive abilities depends on the economy's information structure and on whether or not we impose a condition of externality-freeness.

Information structure. We assume that it is commonly known that, with probability 1, one individual is high-skilled and one individual is low-skilled.[17] I.e. only the states (ω_l, ω_h) and (ω_h, ω_l) have positive probability, whereas the states (ω_l, ω_l) and (ω_h, ω_h) have probability zero. This implies that any one individual knows the other individual's type: If individual 1 observes that the own productive ability is high, then she can infer that the productive ability of individual 2 is low and vice versa. Put differently, each possible state of the economy gives rise to a complete information type space, with the mechanism designer as the only uninformed party.

The Mirrleesian approach. A Mirrleesian analysis imposes externality-free- ness and anonymity. Externality-freeness requires that the outcome for any one individual depends only on that individual's productive ability and not on the productive ability of the other person. Anonymity requires that these outcomes are identical across individuals, so that e.g., the outcome specified for person 1 in case that $\omega_1 = \omega_l$, equals the outcome specified for person 2 in case that $\omega_2 = \omega_l$. Consequently, a social choice function can be represented by two bundles (y_l, c_l) and (y_h, c_h) so that, for all i,

$$(y_i(\omega), c_i(\omega)) = \begin{cases} (y_l, c_l) & \text{whenever} \quad \omega_i = \omega_l \ , \\ (y_h, c_h) & \text{whenever} \quad \omega_i = \omega_h \ . \end{cases}$$

Incentive compatibility then requires that an individual with low productive ability prefers (y_l, c_l) over (y_h, c_h), and that an individual with high productive ability prefers

[17] This setup is due to Piketty (1993). We investigate the Mirrleesian approach under the same information structure.

(y_h, c_h) over (y_l, c_l).[18] Formally,

$$u(c_l)-v\left(\frac{y_l}{w_l}\right) \geq u(c_h)-v\left(\frac{y_h}{w_l}\right) \quad \text{and} \quad u(c_h)-v\left(\frac{y_h}{w_h}\right) \geq u(c_l)-v\left(\frac{y_l}{w_h}\right) \,. \tag{4.10}$$

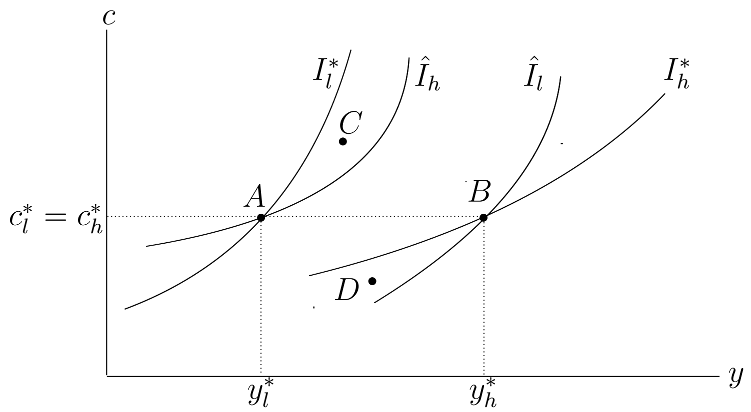
Obviously, these Mirrleesian incentive constraints are violated by the first-best utilitarian welfare maximum. An optimal Mirrleesian allocation is obtained by choosing (c_l, y_l) and (c_h, y_h) so as to maximize the sum of utilities

$$u(c_l) - v\left(\frac{y_l}{\omega_l}\right) + u(c_h) - v\left(\frac{y_h}{w_h}\right) \,,$$

subject to the incentive constraints in (4.10) and the resource constraint $c_l + c_h \leq y_l + y_h$.

Piketty's approach. Piketty (1993) constructs a mechanism which achieves the first-best utilitarian outcome in dominant strategies. This mechanism is anonymous, but not externality-free. The construction is illustrated in Figure 4.1.

Figure 4.1: First-best utilitarian welfare maximum



The Figure illustrates how the first-best utilitarian welfare maximum can be achieved with a mechanism that is not externality-free. I_l^* is the less able individual's indifference curve through $A = (y_l^*, c_l^*)$. Analogously, I_h^* is the more able individual's indifference curve trough $B = (y_h^*, c_h^*)$. The less able individual's indifference curve through B is denoted by $\hat{I}_l$, and the more able individual's indifference curve through A is denoted by $\hat{I}_h$.

In this Figure, point A is the outcome for any one individual if it reports ω_l and the other individual reports ω_h. Point B is the outcome for an individual that reports ω_h if the

[18] According to the Taxation Principle, see Hammond (1979) and Guesnerie (1995), these incentive constraints are equivalent to the possibility to reach a social choice function by specifying a tax schedule $T : y \mapsto T(y)$ so that any one individual i chooses c_i and y_i so as to maximize utility subject to the constraint that $c_i \leq y_i - T(y_i)$.

other individual reports ω_l. Point C is the outcome for an individual that reports ω_h if the other individual also reports ω_h. Analogously, D is the outcome for an individual that reports ω_l if the other individual also reports ω_l. It can easily be verified that truth-telling is a dominant strategy for selfish individuals if (i) point C lies above point A and between the two individuals' indifference curves through A, and (ii) point D lies below point B and between the two individuals' indifference curves through B. Also note that this is incompatible with externality-freeness which would require that $A = D$ and $B = C$.

Social preferences. Models of social preferences can rationalize deviations from this dominant strategy equilibrium. Consider first a model with intentions, such as Rabin (1993). The high-skilled individual might reason in the following way: The other individual could have reported a high ability type, in which case I would have gotten point C. This would have been good for me. So, the other individual is unkind since she did not make use of this possibility to increase my payoff. I am therefore willing to give up own payoff, so as to reciprocally harm the other individual. So, I should declare to be of the low ability type. In this case we both get D. This clearly harms myself and the other person. However, the point D is not that much worse for me, so the possibility to harm the other person is worse the sacrifice. Alternatively, we may consider a model with inequity aversion such as the Fehr-Schmidt-model. In this case, the same deviation could be rationalized by the observation that if both get D, their outcomes are equal, whereas they are very unequal in the dominant strategy equilibrium. Again, if point D is sufficiently close to B achieving this gain in equity is not too costly for an individual with high ability. With the Mirrleesian approach, by contrast, models of social preferences would predict truthful behavior. Since the Mirrleesian mechanism is externality-free, Proposition 4.2 implies that it is social-preference-robust.

An experiment. In the following we report on a laboratory experiment so as to check whether Piketty's approach does indeed provoke more deviations from truth-telling, and, if, yes, what this implies for the levels of utilitarian welfare that are generated by the two mechanisms. The experiment was based on functional form assumptions and parameter choices that are detailed in the following example.

Example 4.4. We impose the following functional form assumption on preferences:

$$U_i = u(c_i) - v\left(\frac{y_i}{\omega_i}\right) = \sqrt{c_i} - \frac{1}{2}\left(\frac{y_i}{w_i}\right)^2 .$$

In addition, we let $\omega_l = 4$ and $\omega_h = 6$. Under these assumptions, the optimal Mirrleesian allocation is given by $(c_l^M, y_l^M) = (3.45, 2.47)$ and $(c_h^M, y_h^M) = (6.23, 7.21)$. The normal form game that is induced by the Mirrleesian mechanism on a complete information type space so that individual 1 is of low ability and individual 2 is of high ability is summarized in the following table. The entries in the matrix are the players' utility levels under the assumption of selfish preferences.

Table 4.6: The game induced by the Mirrleesian mechanism for $(\omega_1, \omega_2) = (\omega_l, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.26, 3.70)$	$(3.26, 3.72)$
ω_h	$(1.99, 3.70)$	$(1.99, 3.72)$

To see that incentive compatibility holds note that first that player 1 does not benefit from claiming to be of high ability if player 2 behaves truthfully. His payoff would drop from 3.26 to 1.99. Analogously, if player 1 behaves truthfully, player 2 does not benefit from understating her ability, her payoff would drop from 3.72 to 3.70. In addition, externality-freeness holds: If player 1 communicates her low type truthfully, then she gets a payoff of 3.26 irrespectively of whether player 2 communicates a high or a low type. Also, if player 2 reveals his high type, he gets 3.72 irrespectively of whether player 1 communicates a high or a low valuation.

Piketty's mechanism is characterized by four points A, B, C and D, as illustrated in Figure 4.2. Points A and B coincide with the first-best utilitarian welfare maximum, so that

$$A = (c_l^*, y_l^*) = (5.53, 3.40) \quad \text{and} \quad B = (c_h^*, y_h^*) = (5.53, 7.66) \; .$$

There is a degree of freedom for the location of the points C and D. To have a completely specified example we need to determine these points in a specific way. We do this so as to capture the desire for welfare-maximizing redistribution which is the basic premise of an analysis of optimal income tax systems. In particular, suppose that there is a small probability, possibly zero, that both types have low abilities. In this case truth-telling of both individuals yields point D. Also suppose that there is an equally small probability that both types have high abilities, which would yield point C. We now allow for the possibility to redistribute resources away from the lucky state in which everybody is of high ability to the unlucky state in which everybody is of low ability. Moreover, we maximize this level of redistribution subject to the constraint of satisfying the principles of Piketty's construction. More formally, we choose point $C = (y^C, c^C)$ so that we extract a maximal tax payment subject to the constraint that C lies above point A and between the two relevant indifference curves through A.[19] We then choose point $D =$

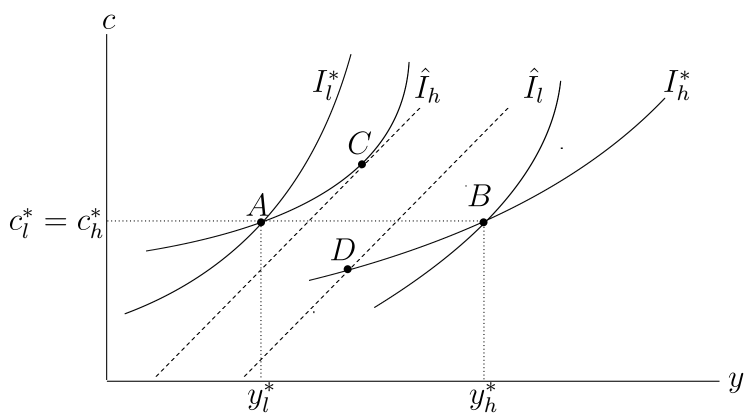
[19]Formally, it is obtained as a solution to the following problem: Maximize $y^C - c^C$, s.t.

$$u(c^C) - v\left(\frac{y^C}{w_h}\right) \geq u(c^A) - v\left(\frac{y^A}{w_h}\right) \quad \text{and} \quad u(c^C) - v\left(\frac{y^C}{w_l}\right) \leq u(c^A) - v\left(\frac{y^A}{w_l}\right) \; .$$

(y^D, c^D) so as to maximize $u(c^D) - v\left(\frac{y^D}{w_l}\right)$ subject to the constraint that $c^D - y^D = y^C - c^C$ and subject to the requirement that point D lies below point B and between the two relevant indifference curves through B. This construction is illustrated in Figure 4.2. It yields the following numerical values

$$C = (y^C, c^C) = (6.45, 7.77) \quad \text{and} \quad D = (y^D, c^D) = (4.62, 3.30) \ .$$

Figure 4.2: Piketty mechanism



The Figure illustrates a specific Piketty mechanism. Point C is chosen so as to extract maximal tax revenues which yields the tangency condition that is shown in the Figure. These tax revenues are then used to make point D as attractive as possible, so that D is determined by the intersection of indifference curve I_h^* and a line with slope 1 and intercept $y^C - c^C$.

The normal form game that is induced by this version of a Piketty mechanism on a complete information type space, so that individual 1 is of low ability and individual 2 is of high ability, is summarized in Table 4.7. Again, the entries in the matrix are the players utility levels under the assumption of selfish preferences.

Table 4.7: The game induced by the Piketty mechanism in Figure 4.2 for $(\omega_1, \omega_2) = (\omega_l, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(2.32, 3.06)$	$(3.98, 3.08)$
ω_h	$(1.04, 4.38)$	$(2.94, 4.40)$

Truth-telling is a dominant strategy equilibrium under the assumption of selfish preferences. Externality-freeness is violated: If player 1 truthfully communicates a low ability type, his payoff depends on what player 2 communicates. Likewise, if player 2 communicates her high ability type truthfully, then her payoff depends on the type declared by player 1.

We conducted two laboratory treatments, one for the Mirrleesian approach and one for Piketty's approach.[20] As expected, we find hardly any deviations from truth-telling with the Mirrleesian approach: 124 of 126 low skilled individuals and 122 of 126 high skilled individuals truthfully report their ability. With Piketty's approach we also find almost no deviations from truth-telling for low skilled individuals, 121 of 126 reports are truthful. This changes with high skilled individuals in Piketty's approach where we observed 21 of 126 individuals to understate their skill level. This is a significantly larger proportion of deviations than with the Mirrleesian approach (two-sided Fisher's exact test, p<0.001). As a result, the Mirrleesian approach reaches an average welfare level that is with 6.93 significantly larger than the average welfare level of 6.78 which results from Piketty's approach (two-sided t-test, $p = 0.014$).

4.9 Concluding remarks

This paper shows how social preferences can be taken into account in robust mechanism design. We have first characterized optimal mechanisms for a bilateral trade problem and a problem of redistributive income taxation under selfish preferences. We have argued theoretically that such a mechanism will not generally produce the desired behavior if individuals have social preferences, and we have illustrated in a laboratory experiment that deviations from the intended behavior indeed occur. We have then introduced an additional constraint on mechanism design, which we termed externality-freeness. We have shown theoretically that such a mechanism does generate the intended behavior if individuals are motivated by social preferences, without a need to specify (beliefs about) the nature and intensity of social preferences. We have finally confirmed in a series of experiments, taking other assumptions in mechanism design for granted (see below), that an externality-free mechanism does indeed generate the intended behavior.

[20] In Piketty's approach only states with a low and a high skilled individual have a positive probability in theory. Despite this, we asked subjects to report actions for all four skill combinations in order to use the same procedures as in our other treatments. The results reported in this section are based on the two states with positive probability. The full experiment data can be found in Appendix 4.D.

These observations raise the question whether externality-freeness is a desirable property. In our analysis of the bilateral trade problem mechanisms were designed with an objective of profit-maximization. In our experiment data, profits were higher with the mechanism that was derived on the assumption that individuals are selfish. The externality-free mechanism would have been more profitable only if more individuals had deviated from truth-telling. This shows, the ability to make money is not necessarily the same as the ability to predict behavior. For a problem of welfare-maximizing income taxation, by contrast, an externality-free Mirrleesian income tax system not only controlled behavior better but also generated more welfare than an alternative mechanism, proposed by Piketty (1993), that would give rise to first-best outcomes if all individuals were selfish.

These observations reflect that externality-freeness is a sufficient but not a necessary condition for the ability to predict behavior. Its advantage is that it successfully controls the underlying motivations across a wide variety of social preferences discussed in the literature, as well as the frequently observed large heterogeneity in parameter values across individuals. A fully fledged mechanism design exercise would require to elicit not only the monetary payoffs of individuals but also the precise functional form of their social preferences. We conjecture that with such a more fine-tuned mechanism design approach, there would no longer be a tension between the ability to predict behavior and the ability to reach a given objective. However, a need to specify the details of the nature and intensity of social preferences, which typically differ across individuals and contexts, would work against our goal to develop robust mechanisms in the spirit of the Wilson doctrine. We leave the question what can and what cannot be reached with a fine-tuned approach to future research.

As an alternative to such an axiomatic approach one might simply try to identify the relevant deviations from selfish behavior empirically, e.g., with a laboratory experiment, and then impose externality-freeness conditions only locally so as to eliminate the specific deviations that pose problems for the mechanism design problem at hand. This approach has the advantage that it does not impose as many additional constraints on the mechanism design problem. The disadvantage is that it does not eliminate all the deviations from selfish behavior that can be rationalized by models of social preferences. Thus, it is not as robust as an externality-free mechanism. We demonstrated the attractiveness of such an engineering approach in the context of the bilateral trade problem. Imposing externality-freeness only locally enabled us to find a mechanism that outperformed both an optimal mechanism for selfish agents and an optimal externality-free mechanism.

Adding behavioral aspects to the mechanism design literature is a promising line of research. That said, we caution that our study cannot, of course, capture all behavioral aspects that seem relevant. For instance, our experiments do not shed light on social preference robustness with incomplete information about monetary payoffs. As a first step, we rather take the theoretically predicted equivalence of implementability in all complete information environments and implementability in all incomplete information environments, as well as the revelation principle, for granted. This way, we can focus on the role of social preferences under certainty in mechanism design, abstracting away from other potential influencing behavioral factors which may arise in cognitively and socially more demanding environments. For instance, recent evidence and theory suggest that some patterns of risk-taking in social context are not easily explained by either standard models of decision making under uncertainty nor standard models of social preferences (e.g., Bohnet, Hermann, and Zeckhauser, 2008, Bolton, Ockenfels, and Stauf, 2015, Saito, 2013, Ockenfels, Sliwka, and Werner, 2014). The implications of such findings for robust mechanism design need further attention. By the same token, our approach leaves open the question whether we can generate the behavior that is needed to implement a given social choice function also with an indirect mechanism, which may be empirically more plausible, than a direct revelation mechanism, e.g. one that simply asks individuals whether they are willing to trade at particular prices. These are fundamental questions, and their answers likely generate more important insights on how motivational and cognitive forces affect the behavioral effectiveness and efficiency of economic mechanisms. We are planning to check robustness along those lines in separate studies.

Appendix 4.A Other models of social preferences

In the body of the text, we have shown that the model of Fehr and Schmidt (1999) predicts deviations from truth-telling in certain situations (see *Observation 4.1*). Below, we present analogous findings for two other models of social preferences, Rabin (1993) and Falk and Fischbacher (2006). The Rabin (1993)-model is an example of intention-based social preferences, as opposed to the outcome-based model of Fehr and Schmidt (1999). The model by Falk and Fischbacher (2006) is a hybrid that combines considerations that are outcome-based with considerations that are intention-based. We show that these models also satisfy *Assumption 4.1*, i.e. selfishness in the absence of externalities.

Similar exercises could be undertaken for other models, such as Charness and Rabin (2002), and Dufwenberg and Kirchsteiger (2004). Whether or not these models would predict deviations from truth-telling under the optimal mechanism for selfish agents depends on the values of specific parameters in these models. To avoid a lengthy exposition, we do not present these details here. The preferences in Charness and Rabin (2002), and Dufwenberg and Kirchsteiger (2004) do, however satisfy the assumption of selfishness in the absence of externalities (*Assumption 4.1*).

Rabin (1993). The utility function of any one player i utility takes the form in (4.6). Rabin models the kindness terms in this expression in a particular way. Kindness intended by i towards j is the difference between j's actual material payoff and an equitable reference payoff,

$$\kappa_i(r_i, r_i^b, r_i^{bb}) = \pi_j(r_i, r_i^b) - \pi_j^{e_i}(r_i^b).$$

The equitable payoff $\pi_j^{e_i}(r_i^b)$ is to be interpreted as a norm, or a payoff that j deserves from i's perspective. According to Rabin (1993), this reference point is the average of the best and the worst player i could do to player j, i.e.

$$\pi_j^{e_i}(r_i^b) = \frac{1}{2}\left(max_{r_i \in E_{ij}(r_i)} \pi_j(\theta_j, f(r_i, r_i^b)) + min_{r_i \in E_{ij}(r_i)} \pi_j(\theta_j, f(r_i, r_i^b))\right),$$

where $E_{ij}(r_i)$ is the set of Pareto-efficient reports: A report r_i belongs to $E_{ij}(r_i)$ if and only if there is no alternative report r_i' so that $\pi_i(r_i', r_i^b) \geq \pi_i(r_i, r_i^b)$ and $\pi_j(r_i', r_i^b) \geq \pi_j(r_i, r_i^b)$, with at least one inequality being strict. Rabin models the beliefs of player i about the kindness intended by j in a symmetric way. Thus,

$$\kappa_j(r_i^b, r_i^{bb}) = \pi_i(r_i^b, r_i^{bb}) - \pi_i^{e_j}(r_i^{bb}).$$

Observation 4.3. *Let f be a social choice function that solves a problem of optimal robust mechanism design as defined in Section 4.4.1. Consider a complete information types space for state $(\overline{\theta}_b, \underline{\theta}_s)$ and suppose that $\theta_b = \overline{\theta}_b$. Suppose that f is such that*

$$\pi_b(\overline{\theta}_b, f(\overline{\theta}_b, \underline{\theta}_s)) = \pi_b(\overline{\theta}_b, f(\underline{\theta}_b, \underline{\theta}_s)) > \pi_b(\overline{\theta}_b, f(\overline{\theta}_b, \overline{\theta}_s)) = \pi_b(\overline{\theta}_b, f(\underline{\theta}_b, \overline{\theta}_s)) \; . \qquad (4.11)$$

Suppose that the buyer's and the seller's first and second order beliefs are as in a truth-telling equilibrium. Also suppose that the buyer has Rabin (1993)-preferences with $y_b \neq 0$. *Then the buyer's best response is to truthfully reveal his valuation.*

The social choice function in Example 4.1 fulfills Condition (4.11). Consider Table 4.3. The buyer's incentive constraint binds. Moreover, if the buyer understates his valuation this harms the seller. Since the seller's intention, when truthfully reporting his type, is perceived as kind, the buyer maximizes utility by rewarding the seller. By (4.8), the buyer will therefore announce his type truth-fully *for all* y_b.

Observation 4.4. *Let* f *be a social choice function that solves a problem of optimal robust mechanism design as defined in Section 4.4.1. Consider a complete information types space for state* $(\overline{\theta}_b, \overline{\theta}_s)$ *and suppose that* $\theta_b = \bar{\theta}_b$. *Suppose that* f *is such that (4.11) holds. Suppose that the buyer's and the seller's first and second order beliefs are as in a truth-telling equilibrium. Also suppose that the buyer has Rabin (1993)-preferences with* $y_b \neq 0$. *Then the buyer's best response is to understate his valuation.*

The social choice function in Example 4.1 fulfills Condition (4.11). Consider Table 4.4. We hypothesize that truth-telling is an equilibrium and show that this leads to a contradiction unless the buyer is selfish: The buyer's incentive constraint binds. Moreover, if the buyer understates his valuation this harms the seller. Since the seller's intention, when truthfully reporting his type, is perceived as unkind, the buyer maximizes utility by punishing the seller. By (4.8), the buyer will therefore understate his type *for all* $y_b \neq 0$. Hence, the Rabin model predicts that the buyer will deviate from truth-telling, *for all* $y_b \neq 0$. Put differently, truth-telling is a best response for the buyer only if $y_b = 0$, i.e. only if the buyer is selfish.

Finally, we note that the utility function in the Rabin (1993)-model satisfies Assumption 4.1 for all possible parametrization of the model. The reason is that two actions which have the same implications for the other player generate the same kindness. The one that is better for the own payoff is thus weakly preferred.

Observation 4.5. *Suppose the buyer and the seller have preferences as in (4.6) with parameters* y_b *and* y_s, *respectively. The utility functions* U_b *and* U_s *satisfy Assumption 4.1, for all* $y_b \neq 0$ *and for all* $y_s \neq 0$,

Falk and Fischbacher (2006). We present a version of the Falk-Fischbacher model that is adapted to the two player simultaneous move games that we study. The utility function takes again the general form in (4.6). The kindness intended by player i is now given as

$$\kappa_i(r_i, r_i^b, r_i^{bb}) = \pi_j(r_i, r_i^b,) - \pi_j(r_i^b, r_i^{bb}) ,$$

Moreover, $\kappa_j(r_i^b, r_i^{bb})$ is modeled by Falk and Fischbacher in such a way that

$$\kappa_j(r_i^b, r_i^{bb}) \leq 0 , \tag{4.12}$$

whenever $\pi_i(r_i^b, r_i^{bb}) - \pi_j(r_i^b, r_i^{bb}) \leq 0$. More specifically, the following assumptions are imposed:

(a) If $\pi_i(r_i^b, r_i^{bb}) - \pi_j(r_i^b, r_i^{bb}) = 0$, then $\kappa_j(r_i^b, r_i^{bb}) = 0$.

(b) The inequality in (4.12) is strict whenever $\pi_i(r_i^b, r_i^{bb}) - \pi_j(r_i^b, r_i^{bb}) < 0$ and there exists r_j so that $\pi_i(r_j, r_i^{bb}) > \pi_i(r_i^b, r_i^{bb})$.

(c) If $\pi_i(r_i^b, r_i^{bb}) - \pi_j(r_i^b, r_i^{bb}) < 0$ and there is no r_j so that $\pi_i(r_j, r_i^{bb}) > \pi_i(r_i^b, r_i^{bb})$, then $\kappa_j(r_i^b, r_i^{bb})$ may be zero or positive.

The case distinction in (c) is decisive for the predictions of the Falk-Fischbacher model. If $\kappa_j(r_i^b, r_i^{bb}) > 0$, then Observation 4.1 for the Fehr-Schmidt-model also holds for the Falk-Fischbacher model. If, by contrast, $\kappa_j(r_i^b, r_i^{bb}) = 0$, then Observations 4.3 and 4.4 for the Rabin-model also hold for the Falk-Fischbacher model. In any case, the Falk-Fischbacher satisfies Assumption 4.1, the assumption of selfishness in the absence of externalities.

Observation 4.6. *Suppose the buyer and the seller have preferences as in the model of Falk and Fischbacher (2006) with parameters y_b and y_s, respectively. The utility functions U_b and U_s satisfy Assumption 4.1, for all $y_b \neq 0$ and for all $y_s \neq 0$.*

This follows since $\pi_j(r_i, r_i^b) = \pi_j(r_i', r_i^b)$ implies that $\kappa_i(r_i, r_i^b, r_i^{bb}) = \kappa_i(r_i', r_i^b, r_i^{bb})$. Consequently, two actions that yield the same payoff for the other player generate the same value of $\kappa_i(r_i, r_i^b, r_i^{bb})\kappa_j(r_i^b, r_i^{bb})$.

Appendix 4.B Proofs

Proof of Proposition 4.1. The relaxed problem imposes only the buyer's ex post participation and incentive constraints, as well as the constraint that the expected payments to the seller are equal to the expected payments of the buyer, with expectations computed using the designer's subjective beliefs. Thus, the problem is to choose, for every state $(\theta_b, \theta_s) \in \Theta_b \times \Theta_s$, $q^f(\theta_b, \theta_s)$, $p_b^f(\theta_b, \theta_s)$ and $p_s^f(\theta_b, \theta_s)$ so as to maximize

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) \left(p_s^f(\theta_b, \theta_s) - \theta_s k(q^f(\theta_b, \theta_s)) \right)$$

subject to the following constraints: (i) the resource constraint

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_b^f(\theta_b, \theta_s) \geq \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_s^f(\theta_b, \theta_s) \ , \tag{4.13}$$

(ii) the incentive and participation constraints for the buyer that are relevant if the seller is of the high cost type,

$$\underline{\theta}_b \, q^f(\underline{\theta}_b, \overline{\theta}_s) - p_b^f(\underline{\theta}_b, \overline{\theta}_s) \geq 0 \ , \tag{4.14}$$

$$\overline{\theta}_b\, q^f(\overline{\theta}_b, \overline{\theta}_s) - p_b^f(\overline{\theta}_b, \overline{\theta}_s) \geq 0 \ , \tag{4.15}$$

$$\underline{\theta}_b\, q^f(\underline{\theta}_b, \overline{\theta}_s) - p_b^f(\underline{\theta}_b, \overline{\theta}_s) \geq \underline{\theta}_b\, q^f(\overline{\theta}_b, \overline{\theta}_s) - p_b^f(\overline{\theta}_b, \overline{\theta}_s) \ , \tag{4.16}$$

and

$$\overline{\theta}_b\, q^f(\overline{\theta}_b, \overline{\theta}_s) - p_b^f(\overline{\theta}_b, \overline{\theta}_s) \geq \overline{\theta}_b\, q^f(\underline{\theta}_b, \overline{\theta}_s) - p_b^f(\underline{\theta}_b, \overline{\theta}_s) \ , \tag{4.17}$$

and finally (iii) the incentive and participation constraints for the buyer that are relevant if the seller is of the low cost type. These constraints have the same structure as those in (4.14)-(4.17), except that $\overline{\theta}_s$ is everywhere replaced by $\underline{\theta}_s$.

Obviously, the resource constraint will be binding, so that the objective becomes to maximize

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) \left(p_b^f(\theta_b, \theta_s) - \theta_s k(q^f(\theta_b, \theta_s)) \right)$$

subject to the constraints in (ii) and (iii). The solution can be obtained by solving a separate optimization problem for each seller type. Thus, optimality requires that $q^f(\underline{\theta}_b, \overline{\theta}_s)$, $q^f(\overline{\theta}_b, \overline{\theta}_s)$, $p_b^f(\underline{\theta}_b, \overline{\theta}_s)$, and $p_f^f(\overline{\theta}_b, \overline{\theta}_s)$ are chosen so as to maximize

$$\sum_{\Theta_b} g(\theta_b, \overline{\theta}_s) \left(p_b^f(\theta_b, \overline{\theta}_s) - \theta_s k(q^f(\theta_b, \overline{\theta}_s)) \right)$$

subject to the constraints in (ii); likewise $q^f(\underline{\theta}_b, \underline{\theta}_s)$, $q^f(\overline{\theta}_b, \underline{\theta}_s)$, $p_b^f(\underline{\theta}_b, \underline{\theta}_s)$, and $p_f^f(\overline{\theta}_b, \underline{\theta}_s)$ are chosen so as to maximize

$$\sum_{\Theta_b} g(\theta_b, \underline{\theta}_s) \left(p_b^f(\theta_b, \underline{\theta}_s) - \theta_s k(q^f(\theta_b, \underline{\theta}_s)) \right)$$

subject to the constraints in (iii).

The solution to these problems is well-known, see e.g., Bolton and Dewatripont (2005). Thus, at a solution, the high-valuation buyer's incentive constraint and the low-valuation buyer's participation constraints bind and the other constraints are slack. E.g., if $\theta_s = \overline{\theta}_s$, then (4.14) and (4.17) bind, and (4.15) and (4.16) are not binding. The optimal quantities are then obtained by substituting

$$p_b^f(\underline{\theta}_b, \overline{\theta}_s) = \underline{\theta}_b q^f(\underline{\theta}_b, \overline{\theta}_s)$$

and

$$p_b^f(\overline{\theta}_b, \overline{\theta}_s) = \overline{\theta}_b q^f(\overline{\theta}_b, \overline{\theta}_s) - (\overline{\theta}_b - \underline{\theta}_b) q^f(\underline{\theta}_b, \overline{\theta}_s)$$

into the objective function which yields

$$g(\underline{\theta}_b, \overline{\theta}_s) \left(\left(\underline{\theta}_b - \tfrac{g(\overline{\theta}_b, \overline{\theta}_s)}{g(\underline{\theta}_b, \overline{\theta}_s)} (\overline{\theta}_b - \underline{\theta}_b) \right) q^f(\underline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\underline{\theta}_b, \overline{\theta}_s)) \right)$$

$$+ g(\overline{\theta}_b, \overline{\theta}_s) \left(\overline{\theta}_b q^f(\overline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\overline{\theta}_b, \overline{\theta}_s)) \right) \ .$$

Choosing $q^f(\underline{\theta}_b, \overline{\theta}_s)$ and $q^f(\overline{\theta}_b, \overline{\theta}_s)$ to maximize this expression yields the optimality conditions that are stated in Proposition 4.1 in the body of the text. □

Proof of Proposition 4.3. For the relaxed problem of optimal externality-free mechanism design the objective is, again, the maximization of

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) \left(p_s^f(\theta_b, \theta_s) - \theta_s k(q^f(\theta_b, \theta_s)) \right) .$$

The resource constraint in (4.13) is binding at a solution to this problem, so that the objective can be equivalently written as

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) \left(p_b^f(\theta_b, \theta_s) - \theta_s k(q^f(\theta_b, \theta_s)) \right)$$

The constraints are the low valuation buyer's ex post participation constraints,

$$\underline{\theta}_b q^f(\underline{\theta}_b, \underline{\theta}_s) - p_b^f(\underline{\theta}_b, \underline{\theta}_s) \geq 0 , \tag{4.18}$$

and

$$\underline{\theta}_b q^f(\underline{\theta}_b, \overline{\theta}_s) - p_b^f(\underline{\theta}_b, \overline{\theta}_s) \geq 0 ; \tag{4.19}$$

the incentive constraint for a high type buyer who faces a low cost seller,

$$\overline{\theta}_b q^f(\overline{\theta}_b, \underline{\theta}_s) - p_b^f(\overline{\theta}_b, \underline{\theta}_s) \geq \overline{\theta}_b q^f(\underline{\theta}_b, \underline{\theta}_s) - p_b^f(\underline{\theta}_b, \underline{\theta}_s) , \tag{4.20}$$

and the constraint, that the seller must not be able to influence the high valuation buyer's payoff,

$$\overline{\theta}_b q^f(\overline{\theta}_b, \underline{\theta}_s) - p_b^f(\overline{\theta}_b, \underline{\theta}_s) = \overline{\theta}_b q^f(\overline{\theta}_b, \overline{\theta}_s) - p_b^f(\overline{\theta}_b, \overline{\theta}_s) . \tag{4.21}$$

Note first that the constraint in (4.19) has to bind at a solution to this problem. The payment $p_b^f(\underline{\theta}_b, \overline{\theta}_s)$ enters only in this constraint. Hence, if we hypothesize a solution to the optimization problem with slack in (4.19), we can raise $p_b^f(\underline{\theta}_b, \overline{\theta}_s)$ without violating any constraint, thereby arriving at a contradiction to the assumption that the initial situation has been an optimum.

Second, the constraint in (4.18) binds as well. Suppose otherwise, then it is possible to raise $p_b^f(\underline{\theta}_b, \underline{\theta}_s)$ by some small $\varepsilon > 0$, without violating this constraints. If at the same time, $p_b^f(\overline{\theta}_b, \underline{\theta}_s)$ and $p_b^f(\overline{\theta}_b, \overline{\theta}_s)$ are also raised by ε, then also the constraints in (4.20) and (4.21) remain satisfied. These increases of the buyer's payments raise the objective function, again contradicting the assumption that the initial situation has been optimal.

Third, the constraint in (4.20) has to be binding. Otherwise, it would be possible to raise $p_b^f(\overline{\theta}_b, \underline{\theta}_s)$ without violating this constraint. If at the same time, $p_b^f(\overline{\theta}_b, \overline{\theta}_s)$ is raised by ε, then also (4.21) remains satisfied. One more time, this contradicts the assumption that the initial situation has been optimal.

These observations enables to express the buyer's payments as functions of the traded quantities, so that

$$p_b^f(\underline{\theta}_b, \underline{\theta}_s) = \underline{\theta}_b q^f(\underline{\theta}_b, \underline{\theta}_s) \ ,$$

$$p_b^f(\underline{\theta}_b, \overline{\theta}_s) = \underline{\theta}_b q^f(\underline{\theta}_b, \overline{\theta}_s) \ ,$$

$$p_b^f(\overline{\theta}_b, \underline{\theta}_s) = \overline{\theta}_b q^f(\overline{\theta}_b, \underline{\theta}_s) - (\overline{\theta}_b - \underline{\theta}_b) q^f(\underline{\theta}_b, \underline{\theta}_s) \ ,$$

and

$$p_b^f(\overline{\theta}_b, \overline{\theta}_s) = \overline{\theta}_b q^f(\overline{\theta}_b, \overline{\theta}_s) - (\overline{\theta}_b - \underline{\theta}_b) q^f(\underline{\theta}_b, \underline{\theta}_s) \ .$$

Substituting these payments into the objective function yields

$$g(\underline{\theta}_b, \underline{\theta}_s) \left(\left(\underline{\theta}_b - \tfrac{g(\overline{\theta}_b, \underline{\theta}_s) + g(\overline{\theta}_b, \overline{\theta}_s)}{g(\underline{\theta}_b, \underline{\theta}_s)} (\overline{\theta}_b - \underline{\theta}_b) \right) q^f(\underline{\theta}_b, \underline{\theta}_s) - \underline{\theta}_s k(q^f(\underline{\theta}_b, \underline{\theta}_s)) \right)$$

$$+ g(\overline{\theta}_b, \underline{\theta}_s)(\overline{\theta}_b q^f(\overline{\theta}_b, \underline{\theta}_s) - \underline{\theta}_s k(q^f(\overline{\theta}_b, \underline{\theta}_s)))$$

$$+ g(\underline{\theta}_b, \overline{\theta}_s)(\underline{\theta}_b q^f(\underline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\underline{\theta}_b, \overline{\theta}_s)))$$

$$+ g(\overline{\theta}_b, \overline{\theta}_s)(\overline{\theta}_b q^f(\overline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\overline{\theta}_b, \overline{\theta}_s)))) \ .$$

Choosing $q^f(\underline{\theta}_b, \underline{\theta}_s)$, $q^f(\overline{\theta}_b, \underline{\theta}_s)$, $q^f(\underline{\theta}_b, \overline{\theta}_s)$ and $q^f(\overline{\theta}_b, \overline{\theta}_s)$ so as to maximize this expression yields the optimality conditions stated in Proposition 4.3. □

Proof of Proposition 4.5. We need to show that there is a solution to the optimization problem in Proposition 4.1 that satisfies

$$p_s^f(\underline{\theta}_b, \underline{\theta}_s) - \underline{\theta}_s k(q^f(\underline{\theta}_b, \underline{\theta}_s)) = p_s^f(\overline{\theta}_b, \underline{\theta}_s) - \underline{\theta}_s k(q^f(\overline{\theta}_b, \underline{\theta}_s)) \ ,$$

and

$$p_s^f(\underline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\underline{\theta}_b, \overline{\theta}_s)) = p_s^f(\overline{\theta}_b, \overline{\theta}_s) - \overline{\theta}_s k(q^f(\overline{\theta}_b, \overline{\theta}_s)) \ ,$$

or, equivalently,

$$p_s^f(\underline{\theta}_b, \underline{\theta}_s) - p_s^f(\overline{\theta}_b, \underline{\theta}_s) = \underline{\theta}_s k(q^f(\underline{\theta}_b, \underline{\theta}_s)) - \underline{\theta}_s k(q^f(\overline{\theta}_b, \underline{\theta}_s)) \ , \tag{4.22}$$

and

$$p_s^f(\underline{\theta}_b, \overline{\theta}_s) - p_s^f(\overline{\theta}_b, \overline{\theta}_s) = \overline{\theta}_s k(q^f(\underline{\theta}_b, \overline{\theta}_s)) - \underline{\theta}_s k(q^f(\overline{\theta}_b, \overline{\theta}_s)) \ . \tag{4.23}$$

The right-hand-side of equations (4.22) and (4.23) is pinned down by the characterization in Proposition 4.1. However, this solution leaves degrees of freedom with respect to the specification of the seller's payments. It only requires that the resource constraint binds which implies that

$$\sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_s^f(\theta_b, \theta_s) = \sum_{\Theta_b \times \Theta_s} g(\theta_b, \theta_s) p_b^f(\theta_b, \theta_s) \ . \tag{4.24}$$

Again the right-hand side of this equation is is pinned down by the characterization in Proposition 4.1. Thus, the four payments to the seller $p_s^f(\underline{\theta}_b, \underline{\theta}_s)$, $p_s^f(\overline{\theta}_b, \underline{\theta}_s)$, $p_s^f(\underline{\theta}_b, \overline{\theta}_s)$ and $p_s^f(\overline{\theta}_b, \overline{\theta}_s)$ need to satisfy the three linear equations in (4.22), (4.23) and (4.24). Obviously, there will be more than one combination of payments to the seller that satisfy all of these conditions. □

Appendix 4.C Instructions

The instructions are a translation of the German instructions used in the experiment, and are identical for all participants. The original instructions are available upon request.

Instructions — General Part

Welcome to the experiment!

You can earn money in this experiment. How much you will earn, depends on your decisions and the decisions of another anonymous participant, who is matched with you. Independent of the decisions made during the experiment you will receive 7.00 € as a lump sum payment. At the end of the experiment, positive and negative amounts earned will be added to or subtracted from these 7.00 €. The resulting total will be paid out in cash at the end of the experiment. All payments will be treated confidentially.

All decisions made during the experiment are anonymous.

From now on, please do not communicate with other participants. If you have any questions now or during the experiment, please raise your hand. We will then come to you and answer your question.

Please switch off your mobile phone during the experiment. Documents (such as books, lecture notes etc.) that do not deal with the experiment are not allowed. In case of violation of these rules you can be excluded from the experiment and all payments.

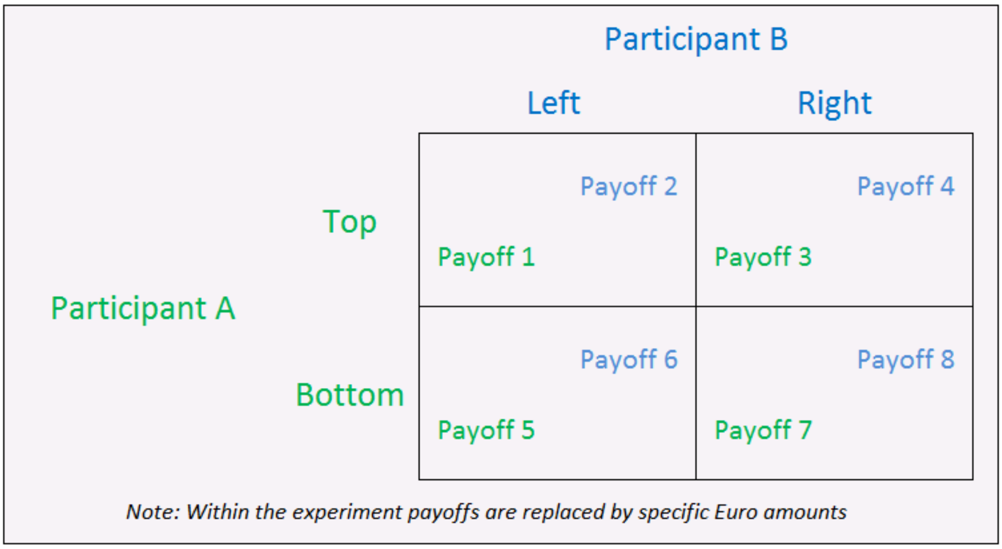
On the following page you will find the instructions concerning the course of the experiment. After reading these, we ask you to wait at your seat until the experiment starts.

First Part — Presentation of decision settings, reading of payoffs

The purpose of this part of the experiment is to familiarize all participants with the decision settings. This ensures that every participant understands the presentation of the decision settings and can correctly infer the resulting payoffs of specific decision combinations. None of the choices in the first part are payoff-relevant.

In the course of this part, eight different decision settings will be presented to you. In all of them two participants have to make a decision without knowing the decision made by the other participant. The combination of the decisions determines the payoffs of both participants. *[These eight decision settings refer to the four complete information games of the respective social choice function of their specific treatment. Each game was presented* *twice: First in the original form and then in a strategically identical form where the payoffs of Participant A and B were switched. This explanation is, of course, not part of the original instructions.]*

Figure 4.3: Exemplary Decision Setting



Participant A, highlighted in green, can decide between *Top* or *Bottom*. Participant B, highlighted in blue, can decide between *Left* and *Right*. The decision of Participant A determines whether the payment results from the upper or lower row in the table. Accordingly, the decision of Participant B determines whether the payment results from the left or right column. Both decisions combined unambiguously determine the cell of the payoff pair.

Each cell contains a payoff pair for both participants. Which payoff is relevant for which participant, is highlighted through their respective color. The green value, which can be found in the lower left corner of every cell, shows the payoff for Participant A. The blue value, which can be found in the upper right corner of every cell, shows the payoff for Participant B.
Please familiarize yourself with the payoff table. Put yourself in the position of both participants and consider possible decisions each participant would make. After a short time for consideration, you can enter a choice combination. The entry can be modified and different constellations can be tried. After choosing two decisions, please enter the payoffs which would result from this constellation. Your entry will then be verified. If your entry is wrong, you will be notified and asked to correct it.

Second Part — Decision Making

At the beginning of the second part you will be assigned to a role which remains constant over the course of the experiment. It will be the role of either Participant A or Participant B. Which role you are assigned to, will be clearly marked on your screen. Please note that the assignment is random, both roles are equally likely. It will be assured that half of the participants are assigned to the role of Participant A and the other half to the role of Participant B.
Simultaneously to the assignment of roles, you are matched with a participant of a different role. This matching is also random. In the course of the remaining experiment you will interact with this participant.

The second part of the experiment consists of four decisions settings. Exactly one decision setting is payoff relevant for you and the other participant matched with you. Which decision setting that is, is determined by chance: Every decision setting has the same chance of being chosen. Hence, please bear in mind that each of the following decision settings can be payoff-relevant.

All decision settings are presented similarly to those of the first part. The difference with respect to the first part is, that you can only make one decision, namely that for your role. Thus, you do not know the decision of the participant matched with you.

Only after you have made a decision for each of the four settings, you will learn which decision setting is relevant for your payoff and the payoff of the participant assigned to you. In addition you will learn the decisions of the other participant in all decisions settings.

After the resulting payoffs are displayed, the experiment ends. A short questionnaire will appear on your screen while the experimenters prepare the payments. Please fill out this questionnaire and wait at your seat until your number is called.

If you have any questions, please raise your hand.

Thank you for participating in this experiment!

Appendix 4.D Supplementary material

4.D.1 The experiment reported on in Section 4.7

Table 4.1”: The game induced by f'' for $(\theta_b, \theta_s) = (\underline{\theta}_b, \underline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(2.68, 6.09)$	$(2.68, 4.05)$
$\overline{\theta}_b$	$(0.97, 6.09)$	$(2.66, 4.86)$

Table 4.2”: The game induced by f'' for $(\theta_b, \theta_s) = (\underline{\theta}_b, \overline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(2.68, 2.65)$	$(2.68, 3.73)$
$\overline{\theta}_b$	$(0.97, -5.79)$	$(2.66, 3.73)$

Table 4.3”: The game induced by f'' for $(\theta_b, \theta_s) = (\overline{\theta}_b, \underline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.41, 6.09)$	$(3.24, 4.05)$
$\overline{\theta}_b$	$(3.43, 6.09)$	$(3.43, 4.86)$

Table 4.4”: The game induced by f'' for $(\theta_b, \theta_s) = (\overline{\theta}_b, \overline{\theta}_s)$.

(π_b, π_s)	$\underline{\theta}_s$	$\overline{\theta}_s$
$\underline{\theta}_b$	$(3.41, 2.65)$	$(3.24, 3.73)$
$\overline{\theta}_b$	$(3.43, -5.79)$	$(3.43, 3.73)$

4.D.2 Choice data T3

Table 4.8: Choice data T3

		Buyer		Seller	
	Game induced by	$\underline{\theta}_b$	$\overline{\theta}_b$	$\underline{\theta}_s$	$\overline{\theta}_s$
T3 *locally externality-free mechanism*	f'' for $(\underline{\theta}_b, \underline{\theta}_s)$	63	0	62	1
	f'' for $(\underline{\theta}_b, \overline{\theta}_s)$	63	0	0	63
	f'' for $(\overline{\theta}_b, \underline{\theta}_s)$	1	62	63	0
	f'' for $(\overline{\theta}_b, \overline{\theta}_s)$	7	56	0	63

4.D.3 Normal form games which are induced by the Mirrleesian mechanism

The game induced by the Mirrleesian mechanism for $(\omega_1, \omega_2) = (\omega_l, \omega_l)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.26, 3.26)$	$(3.26, 1.99)$
ω_h	$(1.99, 3.26)$	$(1.99, 1.99)$

The game induced by the Mirrleesian mechanism for $(\omega_1, \omega_2) = (\omega_l, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.26, 3.70)$	$(3.26, 3.72)$
ω_h	$(1.99, 3.70)$	$(1.99, 3.72)$

The game induced by the Mirrleesian mechanism for $(\omega_1, \omega_2) = (\omega_h, \omega_l)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.70, 3.26)$	$(3.70, 1.99)$
ω_h	$(3.72, 3.26)$	$(3.72, 1.99)$

The game induced by the Mirrleesian mechanism for $(\omega_1, \omega_2) = (\omega_h, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.70, 3.70)$	$(3.70, 3.72)$
ω_h	$(3.72, 3.70)$	$(3.72, 3.72)$

4.D.4 Normal form games which are induced by the Piketty mechanism

The game induced by the Piketty mechanism for $(\omega_1, \omega_2) = (\omega_l, \omega_l)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(2.32, 2.32)$	$(3.98, 1.04)$
ω_h	$(1.04, 3.98)$	$(2.94, 2.94)$

The game induced by the Piketty mechanism for $(\omega_1, \omega_2) = (\omega_l, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(2.32, 3.06)$	$(3.98, 3.08)$
ω_h	$(1.04, 4.38)$	$(2.94, 4.40)$

The game induced by the Piketty mechanism for $(\omega_1, \omega_2) = (\omega_h, \omega_l)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.06, 2.32)$	$(4.38, 1.04)$
ω_h	$(3.08, 3.98)$	$(4.40, 2.94)$

The game induced by the Piketty mechanism for $(\omega_1, \omega_2) = (\omega_h, \omega_h)$.

(U_1, U_2)	ω_l	ω_h
ω_l	$(3.06, 3.06)$	$(4.38, 3.08)$
ω_h	$(3.08, 4.38)$	$(4.40, 4.40)$

4.D.5 Choice data T4 and T5

Table 4.9: Choice data T4 and T5

	Game induced by	*Individual* 1 ω_l^1	ω_h^1	*Individual* 2 ω_l^2	ω_h^2
T4 *Mirrleesian approach*	(ω_l, ω_l)	62	1	62	1
	(ω_l, ω_h)	62	1	2	61
	(ω_h, ω_l)	2	61	62	1
	(ω_h, ω_h)	2	61	2	61
T5 *Piketty's approach*	(ω_l, ω_l)	57	6	55	8
	(ω_l, ω_h)	60	3	14	49
	(ω_h, ω_l)	7	56	61	2
	(ω_h, ω_h)	2	61	9	54

Bibliography

Acharya, V. et al. (2010). "Measuring Systemic Risk". AFA Denver Meetings Paper.

Alesina, A. and F. Passarelli (2014). "Regulation versus Taxation". *Journal of Public Economics* 110, 147–156.

Ariely, D., A. Ockenfels, and A. Roth (2005). "An Experimental Analysis of Ending Rules in Internet Auctions". *RAND Journal of Economics* 36, 890–907.

Arrow, K. (1979). "The Property Rights Doctrine and Demand Revelation under Incomplete Information". In: *Economics and Human Welfare*. Ed. by M. J. Boskin. New York: Academic Press.

Baron, D. and R. Myerson (1982). "Regulating a Monopolist with Unknown Costs". *Econometrica* 50, 911–930.

Bartling, B. and N. Netzer (2014). "An Externality-Robust Auction: Theory and Experimental Evidence". CESifo Working Paper No. 4771.

Becker, G. S. (1974). "A Theory of Social Interactions". *Journal of Political Economy* 82, 1063–1093.

Bergemann, D. and S. Morris (2005). "Robust Mechanism Design". *Econometrica* 73, 1771–1813.

Bierbrauer, F. (2011a). "Incomplete Contracts and Excludable Public Goods". *Journal of Public Economics* 95, 553–569.

——— (2011b). "On the Optimality of Optimal Income Taxation". *Journal of Economic Theory* 146, 2105–2116.

Bierbrauer, F. and N. Netzer (2012). "Mechanism Design and Intentions". University of Zurich, Department of Economics, Working Paper No. 66.

Bierbrauer, F. et al. (2015). "Robust Mechanism Design and Social Preferences". CES ifo Working Paper No. 4713.

Bohnet, I., G. Hermann, and R. Zeckhauser (2008). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States". *American Economic Review* 98, 249–310.

Bolton, G. and A. Ockenfels (2010). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment". *American Economic Review* 100, 628–633.

Bolton, G., B. Greiner, and A. Ockenfels (2013). "Engineering Trust - Reciprocity in the Production of Reputation Information". *Management Science* 59, 265–285.

Bolton, G. and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition". *American Economic Review* 90, 166–193.

——— (2012). "Behavioral Economic Engineering". *Journal of Economic Psychology* 33, 665–676.

Bolton, G., A. Ockenfels, and J. Stauf (2015). "Social Responsibility Promotes Conservative Risk Behavior". *European Economic Review* 74, 109–127.

Bolton, P. and M. Dewatripont (2005). *Contract Theory*. Cambridge, MA, MIT Press.

Börgers, T. (2015). *An Introduction to the Theory of Mechanism Design*. Oxford University Press Inc.

Börgers, T. and P. Norman (2009). "A Note on Budget Balance under Interim Participation Constraints: the Case of Independent Types". *Economic Theory* 39, 477–489.

Boyer, M. and J. Laffont (1999). "Toward a Political Theory of the Emergence of Environmental Incentive Regulation". *RAND Journal of Economics* 30, 137–157.

Camerer, C. (2003). *Behavioral Game Theory. Experiments in Strategic Interaction*. Princton Universtiy Press.

Camerer, C. (1995). "Individual Decision Making". In: *Handbook of Experimental Economics*. Ed. by J. Kagel and A. Roth. Princeton University Press. Chap. 8, 587–683.

Charness, A. and M. Rabin (2002). "Understanding Social Preferences with Simple Tests". *Quarterly Journal of Economics* 117, 817–869.

Che, Y. and I. Gale (2006). "Revenue Comparison for Auctions when Bidders have Arbitrary Types". *Theoretical Economics* 1, 95–118.

Chen, Y. (2008). "Incentive-Compatible Mechanisms for Pure Public Goods: A Survey of Experimental Literature". Ed. by C. Plott and V. Smith.

Chen, Y. and T. Sönmez (2006). "Social choice: An experimental study". *Journal of Economic Theory* 127, 202–231.

Chen, Y. et al. (2010). "Social Comparisons and Contributions to Online Communities". *American Economic Review* 100, 1358–1398.

Coase, R. (1960). "The Problem of Social Cost". *Journal of Law and Economics* 3, 1–44.

Cooper, D. and J. Kagel (2013). "Other-Regarding Preferences: A Selective Survey of Experimental Results". Ed. by C. Plott and V. Smith.

Cramton, P., R. Gibbons, and P. Klemperer (1987). "Dissolving a Partnership Efficiently". *Econometrica* 55, 615–632.

Crawford, V. (1985). "Efficient and Durable Decision Rules: A Reformulation". *Econometrica* 50, 817–835.

Crémer, J. and R. McLean (1988). "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions". *Econometrica* 56, 1247–1257.

Crémer, J. and R. McLean (1985). "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent". *Econometrica* 53, 345–361.

Cropper, O. and W. Oates (1992). "Environmental Economics: A Survey". *Journal of Economic Literature* 30, 675–740.

Damania, R. (1999). "Political Competition, Rent Seeking and the Choice of Environmental Policy Instruments". *Environmental and Resource Economics* 13, 415–433.

D'Aspremont, C. and L.-A. Gerard-Varet (1979). "Incentives and Incomplete Information". *Journal of Public Economics* 11, 25–45.

Dierker, E. and H. Haller (1990). "Tax Systems and Direct Mechanisms in Large Finite Economies". *Journal of Economics* 52, 99–116.

Dijkstra, B. (1998). "A Two-Stage Rent-Seeking Contest for Instrument Choice and Revenue Division, Applied to Environmental Policy". *European Journal of Political Economy* 14, 281–301.

Dufwenberg, M. and G. Kirchsteiger (2004). "A Theory of Sequential Reciprocity". *Games and Economic Behavior* 47, 268–298.

Dufwenberg, M. et al. (2011). "Other-Regarding Preferences in General Equilibrium". *Review of Economic Studies* 78, 613–639.

Falk, A. and U. Fischbacher (2006). "A Theory of Reciprocity". *Games and Economic Behavior* 54, 293–315.

Fehr, E. and K. Schmidt (1999). "A Theory Of Fairness, Competition, and Cooperation". *Quarterly Journal of Economics* 114, 817–868.

Feldman, M. and C. Gilles (1985). "An Expository Note on Individual Risk Without Aggregate Uncertainty". *Journal of Economic Theory* 35, 26–32.

Fischbacher, U. (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments". *Experimental Economics* 10, 171–178.

Gershkov, A. et al. (2013). "On the Equivalence of Bayesian and Dominant Strategy Implementation". *Econometrica* 81, 197–220.

Goulder, L. et al. (1999). "The Cost-Effectiveness of Alternative Instruments for Environmental Protection in a Second-Best Setting". *Journal of Public Economics* 70, 329–360.

Greiner, B. (2004). "An Online Recruitment System for Economic Experiments". *Forschung und wissenschaftliches Rechnen* 63, 79–93.

Gresik, T. and M. Satterhwaite (1989). "The Rate at which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms". *Journal of Economic Theory* 48, 304–332.

Grüner, H. and Y. Koriyama (2012). "Public Goods, Participation Constraints, and Democracy: A Possibiliy Theorem". *Games and Economic Behavior* 75, 152–167.

Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation.* Cambridge University Press.

Güth, W. and M. Hellwig (1986). "The Private Supply of a Public Good". *Journal of Economics* Supplement 5, 121–159.

Güth, W. and M. Kocher (2014). “More than Thirty Years of Ultimatum Bargaining Experiments: Motives, Variations, and a Survey of the Recent Literature”. *Journal of Economic Behavior and Organization* 108, 396–409.

Güth, W., R. Schmittberger, and B. Schwarze (1982). “An Experimental Analysis of Ultimatum Bargaining”. *Journal of Economic Behavior and Organization* 3, 367–388.

Hammond, P. (1979). “Straightforward Individual Incentive Compatibility in Large Economies”. *Review of Economic Studies* 46, 263–282.

Harris, M. and A. Raviv (1981). “Allocation Mechanisms and the Design of Auctions”. *Econometrica* 49, 1477–1499.

Harstad, B. (2012). “Climate Contracts: A Game of Emissions, Investments, Negotiations and Renegotiations”. *Review of Economic Studies* 79, 1527–1557.

Hellwig, M. (2003). “Public-Good Provision with Many Participants”. *Review of Economic Studies* 70, 589–614.

——— (2007). “A Contribution to the Theory of Optimal Utilitarian Income Taxation”. *Journal of Public Economics* 91, 1449–1477.

Hoel, M. and L. Karp (2001). “Taxes and Quotas for Stock Pollutants with Multiplicative Uncertainty”. *Journal of Public Economics* 82, 91–114.

Holmström, B. and R. Myerson (1983). “Efficient and Durable Decision Rules with Incomplete Information”. *Econometrica* 51, 1799–1819.

Jehiel, P. and B. Moldovanu (2006). “Allocative and Informational Externalities in Auctions and Related Mechanisms”. In: *Proceedings of the 9th World Congress of the Econometric Society*. Ed. by R. Blundell, W. Newey, and T. Persson.

Judd, K. (1985). “The Law of Large Numbers with a Continuum of I.I.D. Random Variables”. *Journal of Economic Theory* 35, 19–25.

Kagel, J., Y. Lien, and P. Milgrom (2010). “Ascending Prices and Package Bidding: A Theoretical and Experimental Analysis”. *American Economic Journal: Microeconomics* 2, 160–185.

Kagel, J. and A. Roth (2000). “The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment”. *Quaterly Journal of Economics* 115, 201–235.

Keen, M. (2011). “The Taxation and Regulation of Banks”. IMF Working Paper.

Kelly, D. (2005). “Price and Quantity Regulation in General Equilibrium”. *Journal of Econmic Theory* 125, 36–60.

Kittsteiner, T., A. Ockenfels, and N. Trhal (2012). “Heterogeneity and Partership Dissolution Mechanisms: Theory and Lab Evidence”. *Economics Letters* 117, 394–396.

Kos, N. and M. Messner (2013). “Extremeal Incentive Compatible Transfers”. *Journal of Economic Theory* 148, 134–164.

Kosenok, G. and S. Severinov (2008). “Individually Rational, Budget-Balanced Mechanisms and Allocation of Surplus”. *Journal of Economic Theory* 140, 126 –161.

Laffont, J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation.* USA: MIT Press.

Ledyard, J. (1978). "Incentive Compatibility and Incomplete Information". *Journal of Economic Theory* 18, 171–189.

Lewis, T. (1996). "Protecting the Environment when Costs and Benefits are Private". *The RAND Journal of Economics* 27, 819–847.

Mailath, G. and A. Postlewaite (1990). "Asymmetric Bargaining Procedures With Many Agents". *Review of Economic Studies* 57, 351–367.

Mas-Colell, A., M. Whinston, and J. Greene (1995). *Microeconomic Theory.* USA: Oxford University Press.

Matsuo, T. (1989). "On Incentive Compatible, Individually Rational and Ex Post Efficient Mechanisms for Bilateral Trading". *Journal of Economic Theory* 49, 189–194.

McAfee, P. and J. McMillan (1987). "Auctions and Biddings". *Journal of Economic Literature* 25, 699–738.

Mirrlees, J. (1971). "An Exploration in the Theory of Optimum Income Taxation". *Review of Economic Studies* 38, 175–208.

Montero, J. (2002). "Prices versus Quantities with Incomplete Enforcement". *Journal of Public Economics* 85, 435–454.

Mussa, M. and S. Rosen (1978). "Monopoly and Product Quality". *Journal of Economic Theory* 18, 301–317.

Myerson, R. (1981). "Optimal Auction Design". *Mathematics of Operations Research* 6, 58–73.

——— (1985). *Bayesian Equilibrium and Incentive Compatibility: An Introduction.* Cambridge University Press.

Myerson, R. and M. Satterthwaite (1983). "Efficient Mechanisms for Bilateral Trading". *Journal of Economic Theory* 28, 265–281.

Norman, P. (2004). "Efficient Mechanisms for Public Goods with Use Exclusion". *Review of Economic Studies* 71, 1163–1188.

Ockenfels, A., D. Sliwka, and P. Werner (2014). "Bonus Payments and Reference Point Violations". *Management Science* 61, 1496–1513.

Perotti, E. and J. Suarez (2011). "A Pigouvian Approach to Liquidity Regulation". *International Journal of Central Banking* 7, 3–41.

Pigou, A. (1954). *Economics of Welfare.* Vol. 4th edition. Macmillan and Co.

Piketty, T. (1993). "Implementation of First-Best Allocations Via Generalized Tax Schedules". *Journal of Economic Theory* 61, 23–41.

Pizer, W. (2002). "Combining Price and Quantity Controls to Mitigate Global Climate Change". *Journal of Public Economics* 85, 409–434.

Rabin, M. (1993). "Incorporating Fairness Into Game Theory and Economics". *American Economic Review* 83, 1281–1302.

Radner, R. and A. Schotter (1989). "The Sealed-Bid Mechanism: An Experimental Study". *Journal of Economic Theory* 48, 179–220.

Riley, J. and S. W. (1981). "Optimal Auctions". *The American Economic Review* 71, 381–392.

Rob, R. (1989). "Pollution Claim Settlements under Private Information". *Journal of Economic Theory* 47, 307–333.

Roth, A. (2002). "The Economist as Engineer: Game theory, Experimentation, and Computation as Tool for Design Economics." *Econometrica* 70, 1341–1378.

——— (2012). "Experiments in Market Design". Mimeo.

Saito, K. (2013). "Social Preferences Under Risk: Equality of Opportunity vs. Equality of Outcome". *American Economic Review* 7, 3084–3101.

Samuelson, P. (1954). "The Pure Theory of Public Expenditures". *Review of Economics and Statistics* 36, 387–389.

Schmitz, P. (1997). "Monopolistic Provision of Excludable Public Goods under Private Information". *Public Finance/ Finance Publiques* 52, 89–101.

Sobel, J. (2005). "Interdependent Preferences and Reciprocity". *Journal of Economic Literature* 43, 392–436.

Stavins, R. (1996). "Correlated Uncertainty and Policy Instrument Choice". *Journal of Environmental Economic Management* 11, 218–232.

Stein, J. (2012). "Monetary Policy as Financial Stability Regulation". *The Quaterly Journal of Economics* 1, 57–95.

Stiglitz, J. (1982). "Self-Selection and Pareto-Efficient Taxation". *Journal of Public Economics* 17, 213–240.

Valley, K. et al. (2002). "How Communication Improves Efficiency in Bargaining Games". *Games and Economic Behavior* 38, 127–155.

Vickrey, W. (1961). "Counterspeculation, Auctions and Competitive Sealed Tenders". *Journal of Finance* 16, 8–39.

Weitzman, M. (1974). "Prices vs. Quantities". *Review of Economic Studies* 41, 477–491.

Wilson, R. (1987). "Game-Theoretic Analyses of Trading Processes". In: *Advances in Economic Theory.* Cambridge Books Online. Cambridge University Press, 33–70.

Yohe, G. (1978). "Towards a General Comparison of Price and Quantity Controls: A Survey". *Review of Economic Studies* 45, 229–238.

Curriculum Vitae Désirée Rückert

Persönliche Daten

Geburtsdatum:	20.03.1986
Geburtsort:	Konstanz, Deutschland
Staatsangehörigkeit:	deutsch

Universitätsausbildung

seit 08/2011	Promotion im Fach Volkswirtschaftslehre, Universität zu Köln
10/2010-07/2011	Promotion im Fach Volkswirtschaftslehre, Bonn Graduate School of Economics
10/2005-08/2010	Diplom-Volkswirtin, Universität zu Köln

Akademische Positionen

seit 08/2011	Wissenschaftliche Mitarbeiterin am Center for Macroeconomic Research, Universität zu Köln