Essays in Information Economics and Communication

Inauguraldissertation

zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaften

der Universität Mannheim

vorgelegt von

Petra Loerke

im Sommer 2014

Abteilungssprecher:	Prof. Dr. Eckhard Janeba
Referent:	Prof. Dr. Ernst-Ludwig von Thadden
Korreferent:	Prof. Dr. Thomas Tröger

Tag der mündlichen Prüfung:	14. November 2014

Acknowledgments

I would like to thank my supervisors Ernst-Ludwig von Thadden and Thomas Tröger for their valuable advice and insightful comments along these years. They always encouraged me to come up with my own ideas while at the same time provided me guidance to pursue my research. A particular thanks goes to Johannes Hörner for his great support during my time at Yale University and also afterwards. I am especially grateful for his time and effort, our inspiring discussions and his invaluable feedback.

I also wish to express my deepest gratitude to my two co-authors, Mike Felgenahuer and Andras Niedermayer, for their contribution and contagious enthusiasm. I have learned a lot from them and my dissertation has benefited immensely from their motivation and research experience.

I would specially like to thank my fellow graduate students and my colleagues from the Microeconomics group at the Department of Economics for an enjoyable and dynamic working environment, their helpful comments in countless seminars and the guidance they provided throughout the whole research process. Furthermore, I was fortuned to have great friends at the university who I wish to thank for their constant support and the fun we had together. I would also like to thank the Microeconomics group at Cowles Foundation for the fruitful discussions and inspiring seminars, and, in particular, the graduate students at Yale University, who warmly welcomed me and made me have an amazing time. I gratefully acknowledge the financial support from the Deutsche Forschungsgesellschaft that allowed me to focus on my research in all these years.

Finally, yet foremost, I would like to thank my family and my friends for their great personal encouragement. My warmest and whole-hearted thanks go to my parents, Monika and Alexander, and my brother, Roman. Thank you for your patience and support!

Contents

Chapter 1

General Introduction

Ever since Ackerlof has shown the impact of asymmetric information in the market for used cars in his famous lemons' market example (Akerlof, 1970), asymmetric information has been a well-known problem in Economics. It is a fundamental issue that arises in many environments causing inefficiencies and even market breakdowns. Classical examples include insurance markets and principal-agent settings. The importance of understanding asymmetric information was emphasized by the Nobel Prize committee in 2001 when they awarded the Nobel Prize in Economics to Ackerlof, Spence and Stiglitz "for their analyses of markets with asymmetric information". One way to cope with the problem of asymmetric information can be communication and depending on the particular setup different ways of communication are available. This dissertation consists of three self-contained papers, each of them analyzing a different way of communication in an environment with asymmetric information.

In the first chapter, I consider unverifiable communication ("cheap talk"). The players can transmit information by simple chatting but they cannot prove their announcement. It depends on the setup whether players have an incentive to share information and whether the announcements are trustworthy. I analyze a setup in which several players want to provide a public good together. In this case counter acting incentives are present; players try to coordinate to provide the public good but they also try to understate their willingness to pay in order to contribute less. As a further way of communication, I also allow for an independent party who mediates between the players. It is a known theoretical result that such a mediator can sometimes improve the results of communication. I determine all equilibrium outcomes under both ways of communication and compare them to allocations which can be achieved with other mechanisms.

In the second chapter, I consider a setup in which a player cannot lie about his information.

He can be vague or say nothing at all but everything he says can be verified. I study a setup where an agent tries to persuade a principal to take a specific action by revealing information about the state of the world. At the beginning the state is unknown but the agent can acquire information. I show how the rules for persuasion (whether the agent can choose what to reveal or whether he has to reveal all information) affect information acquisition ex ante.

In the third chapter, I study a basic example for asymmetric information; a buyer-seller setting with private information on the side of the seller which can cause a breakdown of the market. To solve the problem of this information asymmetry certification might help. A certifier can observe the quality of a seller and can communicate it to the buyers. However, the certifier maximizes his own profits. Since he is paid by the sellers, he has an incentive to please the sellers and not to reveal bad information to buyers. Then, however, the information of a certifier has no value and as a consequence sellers are not willing to pay a fee for the certification service. The question is how much a certifier should reveal to please the sellers but at the same time to add any value to his information. I apply a model of certification to credit rating agencies. I extend existing results by showing the effect of aggregate uncertainty about the state of the economy on the rating behavior of an agency.

Chapter 1: Communication and Public Goods

Chapter 1 analyzes the effect of communication on the provision of public goods. It studies a situation where players decide about the provision of a public good themselves; e.g. negotiations between nations or agreements between flat mates sharing an apartment. It is well known in Economics that, even though everyone profits from a public good, players try to free ride on the contributions of other players. Additionally, players have private information about their valuations and this private information causes inefficiencies in the provision of public goods. As it rarely happens that players buy a public good together without prior discussion, I include pre-play communication. I consider a voluntary contribution game and add either direct communication between players (cheap talk) or I allow for the help of a neutral party; a mediator. Using mechanism design, I compare equilibrium outcomes of the game with communication to outcomes which can be obtained in any other game. As players are often able to choose whether to participate or not, I consider mechanisms in which players can drop out of the mechanism at any point in time. The main result of this chapter is that the following three sets are equivalent; the equilibrium outcomes with cheap talk, the equilibrium outcomes with mediated communication and the allocation rules that can be implemented with any mechanism in which players can opt out. This yields the conclusions that, first, a mediator can be replaced by direct communication and, second, that any allocation rule that can be implemented with a complex game in which players can opt out, can also be implemented with a voluntary contribution game with cheap talk. Besides its simple structure, one further advantage of the voluntary contribution game with cheap talk is that a designer needs no information about the distribution of types because the rules of the game are independent of the distributions. In an extension I derive the ex-ante efficient allocation rules under voluntary participation at the interim stage and I show that, in the 2-player-case, in all efficient allocation rules no player has to pay more than her own valuation. Combining the results it follows that for two players even if players can drop out only at an interim stage and are committed afterwards, no budget balanced mechanism can increase efficiency in comparison to the voluntary contribution game with cheap talk.

Chapter 2: Bayesian Persuasion with Private Experimentation

Chapter 2 is joint work with Mike Felgenhauer. We study a situation in which an agent tries to persuade a principal to take an action in his favor. However, the optimal decision for the principal depends on the state of the world which is initially unknown to both players. The agent can privately run as many experiments as desired to acquire information about the state and afterwards selectively reveal the results he wants. We assume that the agent can choose the precision of an experiment but that he cannot manipulate the result. We compare the persuasion probability in the sender preferred equilibrium in a setup where the agent observes the outcome privately to a setup in which the principal is able to observe all signals (public experimentation). Public experimentation is studied in Kamenica and Gentzkow (2011). Under private experimentation the agent might search excessively and if he reveals favorable evidence, this evidence should not be taken at face value. The informational content depends on the number of experiments, which in turn is influenced by experimentation costs and the agent's benefit from persuading the principal. We show that the agent prefers to run just one experiment with a sufficiently high precision which deters further private experimentation. In this equilibrium the persuasion probability is lower than in the sender preferred equilibrium under public experimentation. Consequently, the agent is worse off under private experimentation. The principal benefits from the agent's commitment problem not to run further private experiments and prefers private experimentation. We finally show that the persuasion probability decreases and the decision quality increases in the stakes of the agent. This means that the decision maker profits the more the interested party cares about a favorable decision because then the interested party has to provide higher quality information in order to commit not to run several private experiments.

Chapter 3: Crises and Credit Rating Agencies

Chapter 3 is joint work with Andras Niedermayer. This chapter addresses the behavior of a rating agency in the presence of aggregate uncertainty. The quality of a seller's bond is perfectly known to the seller, but unknown to investors. Therefore, a rating agency may serve as a certifier to transmit information about the quality to the buyers. The contribution of this chapter is that it combines idiosyncratic with aggregate uncertainty. Aggregate uncertainty plays a major role in many markets and especially in the light of the recent financial crisis. We investigate the effect of aggregate uncertainty on incentives to distort ratings. We show that a profit maximizing rating agency does not reveal all information and chooses a binary rating: either investment grade or junk bonds. Furthermore, with aggregate uncertainty the cutoff is not at the socially optimal level. Whether the rating agency has an incentive to be too lenient (a negative cutoff) or too strict (a positive cutoff) is determined by three moments of aggregate uncertainty: the mean, the variance and the sum of the third and higher moments. In a period in which the aggregate expected quality has a large variance and low higher order moments, the agency is too lenient. In a period with a high mean it is too strict. These moments can be interpreted as a period before a crisis and it is an empirical question which effect dominates. We provide two extensions of our main result. First, we outline an empirical strategy to determine whether the pro-cyclical or the counter-cyclical effect dominates. We show how the moments of the distribution of aggregate uncertainty can be identified from the prices of financial derivatives. Second, we extend the model to a setup with risk aversion which explains why there can be multiple rating categories (i.e. investment grade, and possibly a second, speculative grade). The reason is that with risk aversion, investors value more precise information about the quality of an asset to reduce risk. We provide numerical examples to illustrate that a hybrid model of risk aversion and aggregate uncertainty preserves the key insights about the rating agency being too lenient or too strict, but additionally predicts multiple rating categories.

Chapter 2

Communication and Public Goods

2.1 Introduction

Consider a situation in which several players have to decide whether or not to buy an indivisible public good. Even though every player profits if the good is provided, they try to avoid paying for the public good themselves and try to free ride on the contributions of other players. This creates the problem of how to finance a public good. In some cases the public good is provided by the government or another public authority which is able to force players to pay for it (e.g. a bridge that is financed through taxes). In those cases, efficient decisions can be made. However, in many situations such an authority does not exist or no authority wants to impose a decision on the players. In this paper we focus on those situations and we analyzes the private provision of public goods. This problem arises in a wide range of situations, from negotiations between nations to agreements between flat mates sharing an apartment. In all cases players must decide on the provision of a public good themselves. If players know each other's valuations for the public good, they can obtain efficient outcomes. However, we are interested in the case where players possess private information about their valuations and this private information is a source of inefficiencies in the private provision of public goods.[1]

The contribution of this paper is twofold. First, we characterize all equilibrium outcomes with pre-play communication in a voluntary contribution game and show that mediation can be replaced by one round of cheap talk. Second, we show that the voluntary contribution

[1] In a setting involving incomplete information, budget balance and voluntary participation, Mailath and Postlewaite (1990) show that there does not exist any game for the provision of a public good that can implement ex post efficient allocations in equilibrium. This extends the result of Myerson and Satterthwaite (1983) for a buyer/seller setting to the provision of public goods.

game with cheap talk is a robust mechanism which can implement all allocation rules that allow for ex-post quitting rights.

We model N players who have to decide whether or not to provide a discrete public good and how to share the costs. The costs are common knowledge but the individual valuations for the good are private information and they are independent across players. The agents play a voluntary contribution game in which the sum of the transfers is equal to the cost if the good is provided and 0 otherwise. This implies that the game has a balanced budget and that there is no money burning.[2] As it rarely happens that players buy a public good together without prior discussion, we include pre-play communication before the voluntary contribution game. First, we consider communication with the help of a mediator. Second, we consider simple non-verifiable direct communication (cheap talk). While sometimes a mediator might not be available, in many settings players still have the possibility to talk directly to each other.

As mentioned before, we consider a situation in which no authority decides on the public good. Therefore, under the private provision of public goods, players are often able to choose whether to participate or not. We depart from the usual assumptions in mechanism design and consider mechanisms in which players can drop out of the mechanism at any point in time. Even at the final stage, where players already know the allocation of a mechanism, they cannot be forced to participate. This implies that a mechanism needs to ensure that in equilibrium every player gets a utility of at least zero. Furthermore, when a player decides which type to announce, she already takes into account that she can drop out if she does not like the result. This changes her incentives to reveal her true type. If all players reveal their true type even though they can drop out at the end, a mechanism is called veto incentive compatible.

The main result of this paper is that all three sets of allocation rules are identical: (i) equilibrium outcomes with cheap talk, (ii) equilibrium outcomes with mediated communication and (iii) allocation rules that can be implemented with any mechanism in which players can opt out. This implies, first, that mediation can be replaced by cheap talk and, second, that any allocation rule that can be implemented with any mechanism in which players can opt out, can also be implemented with a voluntary contribution game with cheap talk. It follows that the voluntary contribution game with cheap talk is a robust mechanism with respect to

[2] Different definitions of a voluntary contribution game exist in the literature. Some papers, alternatively, assume that contributions are lost if the good is not provided and/or if the sum is larger than the costs. The definition of a voluntary contribution game in our paper is close to what some papers call a subscription game.

the distribution of types. Even though players have private information about their own valuation, the mechanism designer needs no information about the distribution of types because the game does not need to be adapted to the distributions.

We derive and present the result in three steps. First, we show that equilibrium outcomes with mediated communication are a subset of allocation rules in which players can drop out. Second, we show that all these allocation rules in which players can drop out can be implemented by the voluntary contribution game with cheap talk. Therefore, in our setting the equilibrium outcomes with mediated communication are a subset of the equilibrium outcomes with cheap talk. Third, we combine this finding with the general result that equilibrium outcomes with cheap talk are a subset of the ones with mediation. Together this implies the equivalence of the three sets.

One important insight to understand the intuition of the result is that in a voluntary contribution game all allocations that a mediator implements are Nash equilibria of the stage game under complete information. Therefore, players are willing to implement these allocations in the game once they know each other's valuations. Direct communication is relevant in two different ways. First, given that there are multiple stage game Nash equilibria for a voluntary contribution game under complete information, communication can serve to coordinate on one equilibrium.[3] Second, under incomplete information we show that there exist equilibria in which it is incentive compatible for players to actually reveal their true valuation. Regarding the equivalence with allocation rules in which players can opt out, it is important that, given the rules of the game, no player has an incentive to bid more than her true valuation. In this way, a player never has to pay more than her valuation and always has a utility of at least 0 independent of her announcement. Thus, a player would never use an option to drop out after any message in the mediated or unmediated communication game. It follows that all equilibrium outcomes are not only incentive compatibility but also veto incentive compatible.

Since veto incentive compatibility is not a generally known incentive constraint and since it is not obvious to check whether an allocation rule satisfies this constraint, we provide sufficient conditions to replace veto incentive compatibility by the more common interim incentive compatibility. Consider allocation rules that are incentive compatible and in addition deterministic, ex post individually rational and satisfy a monotonicity condition. We find that, if an allocation rule is a convex combination of those allocation rules, it is also veto incentive

[3] The proof contains a technical contribution. We show how cheap talk can be used to create a lottery (multidimensional random variable) with arbitrary distribution.

compatible and therefore it is an equilibrium outcome of the voluntary contribution game with communication.

Mechanism design often analyzes mechanisms that satisfy voluntary participation at the interim stage (interim individual rationality). Not all of these allocation rules are contained in the class of veto incentive compatible allocation rules. However, we are interested in whether the efficient ones are. In an extension we derive the ex ante efficient allocation rules under interim individual rationality and we show that, in the 2-player-case, all efficient allocation rules can be ex post individually rational. In other words, the ex ante efficient provision rules under ex post individual rationality are identical to the ones under interim individual rationality which means that ex post individual rationality does not impose any loss of efficiency in comparison to interim individual rationality. We provide an explicit solution for the transfer function and show that it is monotone in players' valuations. Combining the results we can conclude that for two player any efficient allocation rule is veto incentive compatible and an equilibrium outcome of a voluntary contribution game with one round of cheap talk. This implies that even if players can drop out only at an interim stage and are committed afterwards, no budget balanced mechanism can increase efficiency in comparison to the voluntary contribution game with cheap talk.

Related literature

This paper is related to several strands of literature; in particular to those on pre-play communication and on the provision of public goods.

Connected to the literature on cheap talk, which starts with Crawford and Sobel (1982) who study cheap talk in a sender-receiver game where only one player has private information, our work analyzes a setting where not only one side but all players have private information, send messages and take actions. One paper related to cheap talk in a public good problem is Agastya et al. (2007). They construct a setting with two players who can contribute to a joint project. The contributions are sunk costs which gives the players a strong incentive to coordinate. Focusing on the case in which players have a binary message space, they find that players send a message whether they are going to contribute to the project or not and that communication increases efficiency. In our setup, in contrast, the game satisfies budget balance and we consider mediated and unmediated communication for N players and a rich message space.

There exist few papers which also analyze both, mediated and unmediated pre-play communication. In some settings, as we show for the voluntary contribution game, mediation can be replaced by direct communication. One paper closely related to ours is Matthews and Postlewaite (1989). They show that for double auctions with one buyer and one seller equilibrium outcomes with one round of cheap talk are equivalent to equilibrium outcomes with mediated communication.[4] Forges (1999) modifies the double auction for one seller and several buyers and shows that the equivalence between equilibrium outcomes with mediated and unmediated communication only holds for private valuations. However, in other settings even with private valuations mediated communication can have equilibrium outcomes which are not implementable with cheap talk. An example in which only one player sends messages is Goltsman et al. (2009). They include mediated communication in the setup of Crawford and Sobel (1982) and show that depending on the degree of conflict between the parties, mediated communication might perform better than several rounds of cheap talk. Banks and Calvert (1992) analyze cheap talk and mediated communication in a battle of the sexes game with incomplete information where several players can send messages. They find that a mediator may improve efficiency with respect to unmediated communication. Hörner et al. (2011) consider a model of conflict between two parties which have private information about their strength in a fight and they show that a mediator can increase the probability of peace with respect to one round of cheap talk.

These papers, as well as ours, analyze mediated and unmediated communication before a specific game. However, there is also a strand of literature that provides general conditions under which the set of equilibrium outcomes with (several rounds of) cheap talk is equivalent to the set of equilibrium outcomes with a mediator, e.g. Forges (1990), Gerardi (2004) and Forges and Vida (2013). In our game one round of cheap talk with public messages is sufficient to replace a mediator for any number of players and this result is not covered by their work.

Furthermore, this paper also relates to the literature in mechanism design on ex post individual rationality and veto incentive compatibility. The idea how to derive an ex ante efficient outcome in a buyer-seller setting was introduced by Myerson and Satterthwaite (1983). Williams (1987) extends their result by including welfare weights. We apply his approach to a public good problem. Another paper which is closely related to ours is Gresik (1991). He shows that for bilateral trade the ex ante efficient allocation is also ex post individually

[4] They prove that given the realization of all types (ex post) the probability of trade and the expected transfers are the same. Considering stochastic mechanisms there is a difference between their and our definition of equilibrium outcomes. However, we show equivalence at the posteriori stage which implies equivalence at the ex post stage.

rational and we make use of his method to prove that the same is true for all ex ante efficient allocation rules in a two player public good problem.

Matthews and Postlewaite (1989) and Forges (1999) use the idea of a veto constraint to determine the relation between mediated an unmediated communication in their setup. Besides these two papers, Compte and Jehiel (2009) is one of the few papers analyzing mechanisms with veto constraints. They consider a problem where players bargain over a pie and have different outside options which are private information and can be correlated. This problem is equivalent to a public good setting with correlated types in which every player is essential for the provision of the public good. They prove that even if types are correlated there is no veto incentive compatible mechanism which is ex post efficient.[5] Furthermore, they show that for two players, symmetric welfare weights and independent, uniformly distributed outside options the ex ante efficient allocation satisfies veto incentive compatibility. We extend this result to asymmetric welfare weights and distributions with monotone hazard rates.

The remainder of the paper is structured as follows. Section 2 presents the model. Section 3 describes all equilibrium outcomes of the game with mediated and unmediated communication and shows the equivalence result. Section 4 identifies the ex ante efficient allocation rules under interim individual rationality and budget balance and shows that for two players the efficient allocation rules are ex post individually rational. Section 5 concludes.

2.2 Model

2.2.1 Setup

There are N players. Players must decide whether to provide an indivisible public good or not. The public good costs c and it is non-excludable. Every player i has private information about her own valuation v_i for the public good. This valuation is distributed according to F_i with support $v_i \in [0, \overline{v}_i] \equiv V_i$ and values are drawn independently. We assume that $c > 0$ and $\max_i \overline{v}_i < c$. This information is common knowledge.

Players are risk-neutral with quasi-linear utility. Denote the transfer of player i by X_i. The realized utility of player i is $\pi_i(v_i, Z, X_i) = v_i Z - X_i$ with Z being the decision whether the public good is provided, $Z \in \{0, 1\}$.

We now define an extensive form game. The game is a voluntary contribution game with

[5] This result relates to Cremer and McLean (1988) and McAfee and Reny (1992) who show that, if types are correlated, there are incentive compatible mechanisms with voluntary participation at the interim stage that can implement ex post efficient allocation rules.

communication ex ante, i.e., it has two stages: first, a communication stage; second, a bidding stage. We are going to consider either mediated communication or one round of cheap talk at the communication stage.

Under mediated communication players send privately a message to a "mediator" who then confidentially recommends a bid to each player. In the bidding phase, all players simultaneously submit a bid $b_i \in \mathbb{R}^+$. The rules of the game are that the good is provided if and only if $\sum b_i \geq c$.[6] Denote $z(b) = 1$ if the good is provided and 0 otherwise. We require budget balance; if the good is provided, transfers must add up to the costs: $\sum x_i(b_i, b_{-i}) = c$; if not $\sum x_i(b_i, b_{-i}) = 0$. $x_i(b)$ is the transfer of player i depending on the bids of all players. The cost-sharing rule of the game is exogenously given and satisfies:[7]

- If $\sum b_i \geq c$, the good is provided and the costs are shared according to a sharing rule $x(b)$ that satisfies (i) that no player pays more than her bid, $0 \leq x_i(b) \leq b_i$, and (ii) that the transfer is weakly increasing in the own bid; if $b_i > b_i'$, then $x_i(b_i, b_{-i}) \geq x_i(b_i', b_{-i})$ for all b_{-i}.
- If $\sum b_i < c$, $x_i(b) = 0$ for all i.[8]

One common cost sharing rule that satisfies the conditions above is a proportional cost sharing rule

$$x_i(b_i, b_{-i}) = \begin{cases} \frac{b_i}{\sum_{j=1}^N b_j} c & \text{if } \sum_{j=1}^N b_j \geq c \\ 0 & \text{otherwise.} \end{cases}$$

However, the cost sharing rule does not need to be symmetric, e.g., for two players one alternative rule is that, if the good is provided, one player has to pay her bid completely while the other player only has to pay what is left to cover the costs; $x_i(b_i, b_j) = b_i$ and $x_j(b_i, b_j) = c - b_i$ if the good is provided and otherwise both pay 0.

Define the utility of player i in the game as $u_i(v_i, b_i, b_{-i}) = v_i z(b_i, b_{-i}) - x_i(b_i, b_{-i})$. The agent's strategy specifies for each type which message to send to the mediator, $V_i \to \mathbb{R}$, and how much to bid given the type announced and the recommendation of the mediator, $V_i \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}_0^+$.

Under unmediated communication (cheap talk) all players simultaneously send a message

[6] To simplify notation we drop the index i in the sum over all players $\sum_i$.

[7] Jackson and Moulin (1992) allow for the same family of cost-sharing rules. They provide a mechanism that implements first best allocation rules in an environment where players know each other valuations.

[8] Assume for simplicity that every player is committed to pay her transfer $x_i(b)$ (which is weakly less than her bid b_i). We will show later that every player is willing to pay $x_i(b)$ because no player has an incentive to bid more than her true valuation and thus, $x_i(b) \leq v_i$. The subsequent results also hold if we relax this assumption and assume that the good is not provided if at least one player refuses to pay $x_i(b)$.

$m_i \in [0, \overline{v}_i]$ and in addition a number $y_i \in [0, 1]$, which is used for a jointly controlled lottery. Both messages are publicly observable. In the bidding phase, all players simultaneously submit a bid $b_i \in \mathbb{R}^+$ and the rules of the bidding stage are the same as under mediated communication. The strategy for the agent specifies for each type which message and which number to send to the other players, $V_i \to \mathbb{R} \times [0, 1]$, and how much to bid given the own announcement and the announcements of all the other players, $V_i \times \mathbb{R} \times [0, 1] \times \mathbb{R}^{N-1} \times [0, 1]^{N-1} \to \mathbb{R}_0^+$.

The Equilibrium concept in both games is Perfect Bayesian Equilibrium.

2.2.2 Definitions

An *allocation* consists of the decision $Z \in \{0, 1\}$ whether the public good is provided and a vector of transfers X, where X_i is the payment of player i. An allocation is an element of $\{0, 1\} \times \mathbb{R}^N$.

Let $\mathcal{H}$ denote the Borel σ-algebra on $\{0, 1\} \times \mathbb{R}^N$ and $\mathcal{L}$ the set of probability measures on $\mathcal{H}$. An *allocation rule* r is a mapping $r : V^N \to \mathcal{L}$. It determines which allocations are implemented with positive probability for each vector of private valuations $v \equiv (v_1, \cdots, v_N) \equiv (v_i, v_{-i})$, $v \in V^N$, $V^N \equiv [0, \overline{v}_1] \times \cdots \times [0, \overline{v}_N]$.[9]

An allocation rule is an *equilibrium outcome* of the public good game with communication if there exists a Perfect Bayesian equilibrium in which, for every $v \in V^N$, each allocation has the same probability as in the allocation rule.

We define $p(v)$ as the probability that the good is provided and $t(v)$ as the expected transfer given v. In this way we can write the expected utility of player i given her own valuation and the valuation of the other players as $v_i p(v) - t_i(v)$. Given an allocation rule r, define the expected utility of player i if she sends a message being type m_i as $U_i(m_i, v_i) \equiv E_{v_{-i}}[v_i p(m_i, v_{-i}) - t_i(m_i, v_{-i})]$.

An allocation rule is *posteriori individually rational (PostIR)* if $v_i Z - X_i \geq 0$ for all v and all allocations (Z, X) which are implemented with positive probability (see Forges, 1999). PostIR ensures that at the very final stage, once the allocation has realized, every player gets an utility of at least zero. If an allocation rule is PostIR this implies that $Z \sum v_i \geq \sum X_i$ for all implemented allocations. The difference to the more common ex post individual rationality

[9] This encompasses the possibility that an allocation rule randomizes over different allocations, which allows us to include stochastic mechanisms. For a further discussion whether a stochastic mechanism should be defined as a transition kernel or whether it should be represented as a deterministic mechanism which maps into a lottery see Forges (1994).

constraint is that the ex post stage can still hide a final lottery. Ex post individual rationality states that $v_i p(v) - t_i(v) \geq 0$ for all v and all players i. Thus, PostIR implies ex post individual rationality but not the other way round.[10]

An allocation rule is *ex post budget balanced (ExpostBB)* if for all v the sum of transfers $\sum t_i(v)$ is equal to the costs if the public good is provided and 0 otherwise. The idea is that there are no transfers out of or into the mechanism for all realization of valuations.

An allocation rule is *interim incentive compatible (IC)* if $U_i(v_i, v_i) \geq U_i(m_i, v_i)$ for all m_i, v_i. This means that a player's expected utility of a truthful report, given that everybody else reports truthfully as well, is at least as large as her expected utility of announcing any other type.

For the equilibrium analysis we introduce a further incentive constraint, veto incentive compatibility (Veto-IC), which can be motivated as follows.[11] Consider a mechanism that delivers a candidate allocation; imagine that players can drop out or reject the proposed allocation, given some outside option. We normalize the outside option to 0, i.e. that every player can ensure a utility of 0 if she quits the mechanism. Note that PostIR only implies that a player's utility is at least 0 after revealing her type truthfully; not after deviating and announcing another type.[12] Even though the option to opt out might resemble an individual rationality constraint (it ensures a utility of at least 0 on and off path), it affects the truth-telling incentive constraint as well as it curtails the ability of the mechanism to "punish" a player after deviation. Thus, PostIR and IC do not imply Veto-IC.

Example 1: We provide a simple example of an allocation rule that is PostIR and IC but does not satisfy Veto-IC. There are two players, $i \in \{1, 2\}$, and each player has two possible valuations for the public good, $v_i = 2$ or $v_i = 4$. Each type occurs with probability one half. The provision of the public good costs $c = 4$. The provision and transfer rule are as follows:

provision rule	$v_2 = 2$	$v_2 = 4$
$v_1 = 2$	0	1
$v_1 = 4$	1	1

transfer rule	$v_2 = 2$	$v_2 = 4$
$v_1 = 2$	0,0	2,2
$v_1 = 4$	3,1	1,3

It is easy to verify that this allocation rule satisfies PostIR and IC. However, if player 1 is

[10] PostIR and ex post individual rationality are equivalent if a mechanism is deterministic.

[11] Veto-IC was defined by Forges (1999). See also the condition IC^* in Matthews and Postlewaite (1989). Note that Veto-IC is not equivalent to IC^*, even though the motivation is similar. While players can drop out at an posteriori stage in an allocation rule that is Veto-IC, using IC^* players can only ensure a zero utility at the ex post stage.

[12] An interpretation of PostIR is that at the end players should not regret participating in the mechanism.

able to drop out, she prefers to announce type $v_1 = 4$ if she has a true valuation of $v_1 = 2$ and to reject the allocation if player 2 has valuation $v_2 = 2$. Thus, this allocation rule is not Veto-IC.

For a given allocation rule r, let $\tilde{X}(m_i)$ be the distribution over transfers given message m_i and let $\tilde{Z}(m_i)$ be the distribution over the decision Z whether the public good is provided given m_i. Define the expected utility

$$U_i^*(m_i, v_i) \equiv E[\max\{0, \pi_i(v_i, \tilde{Z}(m_i), \tilde{X}(m_i))\}]$$

which is the expected utility of type v_i if she announces being type m_i and rejects an allocation in case it leads to a negative utility.

An allocation rule is *veto incentive compatible (Veto-IC)* if $U_i^*(v_i, v_i) \geq U_i^*(m_i, v_i)$ for all v_i, m_i. In this case every player has an incentive to reveal her true type even though she always has the option to opt out of the mechanism.

2.3 Game with communication

2.3.1 Mediated communication

First, we consider mediated communication in the communication phase. Define a mediator as a communication device $\varphi : \mathbb{R}^N \to \Delta\mathbb{R}^N$ that, after receiving messages from all players, can send messages to all players. This is interpreted as follows: first, each player reports her type to the mediator and then he recommends an action to each player. Using the generalized revelation principle by Myerson (1982, 1985) we restrict attention to equilibria in which every player reveals her type truthfully and follows the recommendation of the mediator.[13]

The following proposition describes all equilibrium outcomes of the game with mediated communication.

[13] The revelation principle states that for any given coordination mechanism φ and for any given Bayesian equilibrium of the induced communication game, there exists a mechanism φ' in which it is incentive compatible for a player to tell her true type and to obey the recommendation and that every type gets the same expected utility as in the equilibrium of the original mechanism φ.

An intuition for this result is that given any mechanism and equilibrium, the mediator can ask the players to reveal their true type, infer the messages they would send and compute the messages he would send to the players. Then he identifies the action the players would take given their true types and the recommendations. Thus, the mediator can tell the players this action directly and they are willing to obey. Therefore, we can restrict to communication mechanisms and equilibria in which the players honestly report their types and follow the recommendation of the mediator.

Proposition 2.1. *Equilibrium outcomes of the mediated communication game satisfy Veto-IC, ExpostBB and PostIR.*

The intuition of the proof is as follows. First, we show that no player has an incentive to bid more than her true valuation and that she is indifferent between bidding her true valuation and bidding more only if for all v_{-i} the higher bid leads to the same transfer x_i as bidding her valuation. Therefore, the equilibrium outcome satisfies PostIR. Second, it follows directly from the rules of the game that all allocation rules must be ExpostBB. In the third part of the proof we use that, even after lying about her type, it is optimal for a player to bid weakly less than her true valuation. Thus, in any resulting allocation each player gets a nonnegative utility and therefore would never use an option to drop out. By incentive compatibility the expected utility of telling the true type and following the recommendation is weakly larger than the utility of deviating to any other message or bid. Since the optimal bidding strategy after any message would never use the possibility to opt out, Veto-IC follows from IC.

2.3.2 Unmediated communication

In this section, we consider non-verifiable face to face communication; one round of cheap talk. To understand the intuition of the equilibrium analysis it helps to look at the stage game under complete information first. Assume every player's valuation is common knowledge and the game only consists of the bidding phase. There exist two classes of Nash equilibria in the stage game with complete information which are of particular interest. Fix $v \in V^N$. In the first class all players bid $b_i = 0$. The public good is not provided. This is an equilibrium because no player can profit by increasing her bid. In the second class fix $\{x_i\}_{i=1}^N$ such that $\sum x_i = c$ and $0 \leq x_i \leq v_i$ for all i. The bid profile $b_i = x_i$ for all i is an equilibrium. Every player pays no more than her valuation. If one player lowers her bid, the good is not provided any longer. Furthermore, no player has an incentive to increase her bid because her transfer is weakly increasing in her own bid. Note that every allocation that is ExpostBB and PostIR is an outcome of a Nash equilibrium of the stage game under complete information.

The following proposition states that Veto-IC, ExpostBB and PostIR are sufficient conditions such that an allocation rule can be implemented as an equilibrium outcomes of the voluntary contribution game with only one round of cheap talk.

Proposition 2.2. *All allocation rules that satisfy Veto-IC, ExpostBB and PostIR are equilibrium outcomes of the public good game with one round of cheap talk.*

The intuition for this result is the following. We show that there always exists an equilibrium in which every player reveals her true type and afterwards the players implement an equilibrium of a stage game with complete information. The announcements y_i create a jointly controlled lottery that serves as a public randomization device. Its realization defines which stage game equilibrium players should play and the lottery is constructed such that the probability for each allocation coincides with the probability of the allocation rule. If a player deviates and reduces her bid, the sum of the bids will be lower than c: the public good is not provided. In the game, a player can either play according to the equilibrium strategy or deviate and get a utility of 0 in case she bids lower. Thus, even though a player has a wide range of deviations, it reduces to a simple choice: follow the equilibrium prescription or to get 0. By definition of Veto-IC, it is optimal for a player to reveal her true type even when she can drop out and ensure a 0 utility given the selected allocation; hence, it is also incentive compatible for the players to reveal their true type in the round of cheap talk.

2.3.3 Equivalence

It is immediate that in general the equilibrium outcomes with cheap talk are a subset of the ones with mediated communication. Instead of sending messages directly to each other (cheap talk), the players can send their messages to a mediator, who just relays them publicly.

$$\boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes}\\ \text{with mediator}\end{array}} \overset{\text{in general}}{\supseteq} \boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes with}\\ \text{cheap talk}\end{array}} \tag{2.1}$$

First, we provide an example by Farrell (1983) to illustrate that the set of equilibrium outcomes with cheap talk can be a strict subset of the ones with mediated communication. That is a mediator can implement equilibrium outcomes that are not possible with cheap talk.

Example 2: There are two players, every player can be one of two types, A or B, and every player choose between two possible actions, X or Y. A mediator recommends both players to play Y if both are type B and otherwise to play X. It is an equilibrium in the game for both players to reveal their true type and follow the recommendation. Thus, the mediator reveals the information whether both are type B or not. Such an information structure cannot be achieved by cheap talk because whether one player reveals her type cannot be conditioned on the announcement of the other player.[14]

[14] A full description of the game is provided in the appendix.

Taken together Proposition 1 and Proposition 2 states that

$$\boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes}\\ \text{with mediator}\end{array}} \overset{\text{Prop.1}}{\subseteq} \boxed{\begin{array}{c}\text{allocation}\\ \text{rules that}\\ \text{satisfy Veto-IC,}\\ \text{ExpostBB}\\ \text{and PostIR}\end{array}} \overset{\text{Prop.2}}{\subseteq} \boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes with}\\ \text{cheap talk}\end{array}} \tag{2.2}$$

which implies that the set of equilibrium outcomes with mediated communication is a subset of the set of equilibrium outcomes with cheap talk. Consequently, combining (2.1) and (2.2), the set of equilibrium outcomes with one round of cheap talk is identical to the set with mediated communication.

$$\boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes}\\ \text{with mediator}\end{array}} \mathrel{\hat{=}} \boxed{\begin{array}{c}\text{allocation}\\ \text{rules that}\\ \text{satisfy Veto-IC,}\\ \text{ExpostBB}\\ \text{and PostIR}\end{array}} \mathrel{\hat{=}} \boxed{\begin{array}{c}\text{equilibrium}\\ \text{outcomes with}\\ \text{cheap talk}\end{array}}$$

This leads to the following proposition.

Proposition 2.3. *The following three sets are equivalent:*

(i) equilibrium outcomes with one round of cheap talk

(ii) equilibrium outcomes with mediated communication

(iii) allocation rules that satisfy Veto-IC, ExpostBB and PostIR

First, this implies that a mediator can be replaced by a single round of cheap talk. Second, it follows that also several rounds of cheap talk cannot implement any allocation rule which cannot be implemented by a single round. Third, there is no other game in which players can drop out that can implement any other allocation rule as an equilibrium outcome than the voluntary contribution game with communication.

Note that the set of equilibrium outcomes is independent of the specific cost sharing rule.[15]

[15] One question is whether our results carry over to related setups, e.g. effort in teams. The main difference is that there contributions naturally are sunk costs. First, consider all equilibrium outcomes under mediated and unmediated communication. All these allocation rules can be implemented by bids which are equal to the required transfers. Thus, they are not affected by a change in the assumption about sunk costs. The general result that the equilibria with cheap talk are a subset of the ones with mediated communication still holds but we do not know whether the sets are identical. Furthermore, equilibrium outcomes do not

2.3.4 Set of veto incentive compatible allocation rules

We have shown that the set of veto incentive compatible allocation rules is equivalent to the set of equilibrium outcomes in a voluntary contribution game with mediated and unmediated communication. Though, Veto-IC is not commonly used and sometimes it is difficult to see whether an allocation rule satisfies this condition. In this section we provide sufficient conditions such that we can replace Veto-IC by the well known interim incentive compatibility (IC). Note that Veto-IC and PostIR imply IC.

We define a monotonicity condition (MON). An allocation rule satisfies MON iff

1. the provision rule is monotone, i.e. $p(v_i, v_{-i})$ is weakly increasing in v_i
2. the transfers are monotone, i.e. $t_i(v_i, v_{-i})$ is weakly increasing in v_i,
3. the transfers satisfy for all $\{v_{-i}|p(v_i, v_{-i}) = 1 \bigcup p(v_i', v_{-i}) = 0\}$ that $t_i(v_i, v_{-i}) \geq \max_{v_k}\{t_i(v_i', v_k)\}$ with $v_i > v_i'$.

Since condition (ii) does not consider for a change in the provision rule, condition (iii) ensures that transfers are high if the provision rule changes from 0 to 1. Note that under ExpostBB for two players the monotonicity condition MON simplifies to the first two conditions. Further, an allocation rule is *deterministic (DET)* if it does not randomize over different allocations given v; $p \in \{0, 1\}$, $p(v) = z(v)$ and $t_i(v) = x_i(v)$.

Proposition 2.4. *An allocation rule that satisfies IC, PostIR, DET and MON also satisfies Veto-IC.*

In the proof we show that no player lies about her true valuation if PostIR, DET and MON are satisfied. The first possible deviation is that a player announces a lower valuation, because a lower type might have lower expected transfers. However, since the allocation rule is PostIR, a player never rejects an allocation after announcing a lower type and therefore Veto-IC follows from IC. The second possible deviation is that a player announces a higher type. Then the expected transfer increases but also the probability that the good is provided increases. It is now provided for v_{-i} for which it is not provided if a player announces her true valuation. If the player does not reject any allocation after announcing a higher type, Veto-IC follows again from IC. The difficult part is to prove that a player has no incentive to lie if she rejects some allocations at the end. Using the monotonicity condition MON, we can

have to be PostIR anymore and thus, also the comparison to Veto-IC mechanisms does not apply. In total, all allocation rules stated in the Proposition are equilibrium outcomes in a game with sunk costs but there might be additional equilibria in which some contributions are lost.

prove this by contradiction. We show that if a player with valuation v_i profits from telling m_i instead of her true valuation, then a higher type v_i', $v_i' \in (v_i, m_i]$, profits even more from announcing m_i. However, a player with type m_i cannot profit from announcing m_i relatively to announcing her true valuation. This yields the contradiction. The intuition is that by requiring monotonicity in the transfers, we ensure that the "gained" realization of v_{-i} (where v_{-i} are relatively low) lead to high transfers for player i. Therefore it does not pay off to announce a higher type.

In order to relax the condition that an allocation rule needs to be DET, we prove that the set of Veto-IC, ExpostBB and PostIR allocation rules is convex.

Lemma 2.1. *The set of allocation rules that satisfy Veto-IC, ExpostBB and PostIR is convex.*

This implies that every convex combination between two allocation rules that satisfy Veto-IC, ExpostBB and PostIR also satisfies these conditions and therefore belongs to the set of equilibrium outcomes.

Combining Proposition 2.2, Proposition 2.4 and Lemma 2.1 gives the following corollary.

Corollary 2.1. *Any allocation rule r that is a convex combination of allocation rules that satisfy IC, PostIR, ExpostBB, DET and MON can be implemented by a public good game with one round of cheap talk.*

2.4 Extension

So far, we have derived the set of equilibrium outcomes of a voluntary contribution game with mediated and unmediated communication. In this section we use mechanism design to analyze the efficiency property of these equilibria within the context of public goods.

In the previous section we have shown that all allocation rules that satisfy Veto-IC, ExpostBB and PostIR are an equilibrium outcome of the game with communication and we provided sufficient conditions for an incentive compatible allocation rule to satisfy Veto-IC. Now we check whether we can relax the posterior participation constraint to a participation constraint at the interim stage. Define that an allocation rule is *interim individual rational (IntIR)* if the expected utility $U_i(v_i, v_i)$ is nonnegative for all v_i. This implies that a player is willing to participate in a mechanism after she knows her own valuation. It is straight forward that not all allocation rules that satisfy IntIR also satisfy PostIR.

First, as a benchmark, we identify the efficient mechanism under interim individual rationality, incentive compatibility and ex post budget balance. Second, we analyze whether the ex ante efficient allocation rules under IntIR satisfy PostIR and whether they also satisfy Veto-IC.

2.4.1 Efficient mechanisms

An ex post efficient allocation rule would provide the public good if the sum of the valuations is larger than the costs, $\sum v_i \geq c$. However, Mailath and Postlewaite (1990) show that there is no ex post efficient mechanism that satisfies IC, ExpostBB and IntIR.[16]

We call an allocation rule *efficient* if it is Pareto-efficient in terms of the players' ex ante expected utilities (ex ante efficient). To find the efficient allocation rule, we maximize the weighted sum of ex ante expected utilities subject to IC, ExpostBB and IntIR. Using the revelation principle, we focus on direct revelation mechanisms which can be characterized by the function $p(v)$, which is the probability that the good is provided, and the transfer function $t(v)$, which gives the expected transfer each player has to pay. From now on we assume that the distributions $F_i(v_i)$ satisfy monotone virtual valuations, i.e., that $v_i - \frac{1-F_i(v_i)}{f_i(v_i)}$ is increasing in v_i. Define the density $f(v) \equiv f_1(v_1)\cdots f_N(v_N)$ and $f_{-i}(v_{-i}) \equiv f_1(v_1)\cdots f_{i-1}(v_{i-1})f_{i+1}(v_{i+1})f_N(v_N)$. Simplify notation in the following way: $\int_0^{\overline{v}_1}\cdots\int_0^{\overline{v}_N}()\,dv_N\cdots dv_1 \equiv \int()\,dv$ and $\int_0^{\overline{v}_1}\cdots\int_0^{\overline{v}_{i-1}}\int_0^{\overline{v}_{i+1}}\cdots\int_0^{\overline{v}_N}()\,dv_N\cdots dv_{i+1}dv_{i-1}\cdots dv_1$ $\equiv \int()\,dv_{-i}$. That is the integral is over the whole support of v (respectively v_{-i}) if no bounds of integration are specified. We define the following expected values: player i's expected payment $\overline{t}_i(v_i) = \int t_i(v_i, v_{-i})f_{-i}(v_{-i})dv_{-i}$, the probability of provision $\overline{p}_i(v_i) = \int p(v_i, v_{-i})f_{-i}(v_{-i})dv_{-i}$ given player i has valuation v_i and player i's expected utility if she reveals her true type $U_i(v_i) = U_i(v_i, v_i) = v_i\overline{p}_i(v_i) - \overline{t}_i(v_i)$. Define furthermore $q_i(v_i, \alpha_i) = v_i - \alpha_i\frac{1-F_i(v_i)}{f_i(v_i)}$ and

$$\begin{aligned}\Gamma(p) &= \int\left(\sum\left(v_i - \frac{1-F_i(v_i)}{f_i(v_i)}\right) - c\right)p(v)f(v)dv \\ &= \int\left(\sum q_i(v_i,1) - c\right)p(v)f(v)dv.\end{aligned}$$

The first proposition is a variation from Theorem 1 in Myerson and Satterthwaite (1983) and it gives necessary and sufficient conditions for an allocation rule to be IC and IntIR.

[16]In fact, Mailath and Postlewaite (1990) show that there is no ex post efficient mechanism that satisfies IC, ex ante budget balance and IntIR. Note that ex ante budget balance is a a weaker condition than ExpostBB.

Proposition 2.5. *(i) If a mechanism is incentive compatible, it holds that*

$$\sum U_i(0) = \int \left[p(v) \left(\sum \left(v_i - \frac{1 - F_i(v_i)}{f_i} \right) - c \right) \right] f(v)dv. \tag{2.3}$$

(ii) There exists a transfer function such that the mechanism is incentive compatible and interim individual rational if and only if $\overline{p}_i(v_i)$ is weakly increasing in v_i and

$$0 \leq \int \left[p(v) \left(\sum \left(v_i - \frac{1 - F_i(v_i)}{f_i} \right) - c \right) \right] f(v)dv. \tag{2.4}$$

Among others, Proposition 2.5 shows that an IC and InterIR allocation rule can be described by a provision rule $p(v)$ and the constraint $\sum U_i(0) = \Gamma(p)$. Incentive compatibility and the provision rule $p(v)$ are sufficient to pin down the expected transfers $\bar{t}_i(v_i)$ up to a constant. This constant is not determined because the mechanism only specifies the sum $\sum U_i(0)$ and not the individual $U_i(0)$ for every player. Note that $U_i(0)$ is the expected utility of a player with the lowest type. Thus, incentive compatibility pins down the expected utility for every type up to a constant and the constant is the utility of the lowest type.

Among all IntIR and ExpostBB mechanisms, we are interested in the ex ante efficient mechanisms. The idea of the proof follows Williams (1987) who characterizes the ex ante efficient allocations in a setting with one buyer and one seller.

Define μ_i as the welfare weight of player i, with $\mu_i \geq 0$, and μ^{max} as the highest welfare weight. We maximize the weighted sum of ex ante expected utilities

$$\sum \mu_i \int_0^{\overline{v}_i} U_i(v_i) f_i(v_i) dv_i \tag{2.5}$$

subject to incentive compatibility and constraint (2.4). Before specifying the efficient mechanisms, we show in the following lemma how the value of $U_i(0)$ depends on the welfare weights.

Lemma 2.2. *Suppose a mechanism $(p, \Gamma(p))$ is an efficient allocation rule. Then $U_i(0) = 0$* for all players with $\mu_i < \mu^{max}$.

The idea is that $\sum U_i(0)$ is given by the mechanism but it is not specified how to distribute it on the individual $U_i(0)$. If all players have equal welfare weights, the ex ante expected utility does not depend on the split of $\sum U_i(0)$. However, if welfare weights are unequal, the ex ante expected utility increases by shifting utility to players with the highest welfare weights. This shift is limited by IntIR because all players must have $U_i(0) \geq 0$.

Knowing how $\sum U_i(0)$ is distributed among the individual players, we can determine the optimal mechanism. For the purpose of the next proposition, we introduce the notation of allocation rules $p^{\alpha}(v)$ which have the following specific form: $p(v) = 1$ if $\sum q_i(v_i, \alpha_i) \geq c$ and 0 otherwise.

Proposition 2.6. *The efficient allocation rule is given by $p^{\alpha}(v)$ which is*

$$p(v) = \begin{cases} 1 & if \sum q_i(v_i, \alpha_i) \geq c, \\ 0 & otherwise. \end{cases}$$

with

$$\alpha_i = \begin{cases} \alpha_i^* \; with \; \alpha_i^* = \frac{\mu^{max} - \mu_i}{\mu^{max}} & if \; \Gamma(p^{\alpha^*}) \geq 0, \\ \frac{\mu^{max} + \lambda - \mu_i}{\mu^{max} + \lambda} \; and \; \lambda \; such \; that \; \Gamma(p^{\alpha}) = 0 & if \; \Gamma(p^{\alpha^*}) < 0. \end{cases}$$

Note that the efficient allocation rules are deterministic and monotone in the provision rule. In comparison to an ex post efficient allocation rule in which $p(v) = 1$ if $\sum v_i \geq c$, the probability that the public good is provided is lower.

Proposition 2.6 distinguishes between allocation rules in which $\Gamma(p) = \sum U_i(0) > 0$ and allocation rules in which $\Gamma(p) = \sum U_i(0) = 0$. The following Lemma shows that in some cases we know that condition $\Gamma(p) = 0$ is binding.

Lemma 2.3. *If the good is never provided for any N-1 players or if all players have equal welfare weights, the condition $\Gamma(p) = 0$ is binding.*

Remember that $U_i(0) = 0$ for all players if $\Gamma(p) = 0$. However, if $\Gamma(p) > 0$, $U_i(0) > 0$ for at least one player. It follows from Lemma 2.3 that for two players $\Gamma(p) = 0$ always binds because we have assumed that one player cannot provide the public good alone, $\max_i \overline{v}_i < c$.

2.4.2 Ex post individual rationality for two players

We restrict to two players and efficient allocation rules. It follows from the previous section that all ex ante efficient allocation rules are deterministic and have a monotone provision rule. The question is whether there exists a transfer function such that the efficient allocation rule also satisfies PostIR and the conditions on the transfers required by MON. Since the efficient allocation rules are deterministic, PostIR is equivalent to ex post individual rationality. To relate to our previous results we use the notation of PostIR in this section. The derivation of the transfer function follows Gresik (1991) who provides a PostIR transfer function in a bilateral trade setting.

In this section we show that there exists such a transfer function if the distributions of types have monotone hazard rates.

Proposition 2.7. *For any efficient mechanism there exists a transfer function that is MON, PostIR and ExpostBB if both hazard rates are monotone.*

To prove this proposition we state an incentive compatible transfer function in Lemma 2.5 and in Lemma 2.6 and 2.7 we show that it is PostIR and monotone if both hazard rates are monotone. Combining Corollary 2.1 with Proposition 2.7 we get the following result.

Corollary 2.2. *For two players any ex ante efficient allocation rule can be implemented by a public good game with one round of cheap talk if the hazard rates are monotone.*

To show Proposition 2.7 define $v_j^*(v_i)$ as the smallest valuation of player j for which the good is provided if player i has valuation v_i. This means that $q_i(v_i) + q_j(v_j^*) = c$.[17] Define $v_i^*(v_j)$ analogously. Assume that $v_i^*(v_j)$ is differentiable and that $\frac{\partial v_j^*(v_i)}{\partial v_i} < 0$ and $\frac{\partial v_i^*(v_j)}{\partial v_j} < 0$. Note that $v_i^*(v_j^*(v_i)) = v_i$ and that $v_i^*(v_j)$ depends on the allocation rule and consequently on the welfare weights. Even though welfare weights will not appear explicitly in the analysis, the results hold for all weights because they are already incorporated in the provision rule. Knowing that $\Gamma(p) = 0$ for all welfare weights leads to $U_i(0) = 0$ for both players and together with the provision rule the expected transfer $\bar{t}_i(v_i)$ is pinned down and we can write them in the following way.

Lemma 2.4. *The expected transfer can be written as*

$$\bar{t}_i(v_i) = \int_{v_j = v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j) f_j(v_j) dv_j. \tag{2.6}$$

The expected transfer of player j can be derived in the same way and written as $\bar{t}_j(v_j) = \int_{v_i = v_i^*(v_j)}^{\overline{v}_i} v_j^*(v_i) f_i(v_i) dv_i$. Budget balance requires that $t_i(v_i v_j) + t_j(v_i, v_j) = c$ if the good is provided and 0 otherwise. Since we are looking for a transfer function that is PostIR, we set $t_i(v) = t_j(v) = 0$ if the good is not provided. Together with budget balance this implies that

[17] Since we restrict to efficient allocation rules, we simplify notation and drop α^* in $q_i(v_i, \alpha^*)$.

$\bar{t}_j(v_j)$ also has to satisfy $\bar{t}_j(v_j) = \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} (c - t_i(v_i, v_j)) f_i(v_i) dv_i$ which leads to

$$\int_{v_i=v_i^*(v_j)}^{\overline{v}_i} v_j^*(v_i) f_i(v_i) dv_i \quad = \quad \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} (c - t_i(v_i, v_j)) f_i(v_i) dv_i. \tag{2.7}$$

Define the following variables:

- $\theta_1(v_j) = -(1 - F_j(v_j)) \left[(v_j + v_i^*(v_j) - c) f_i(v_i^*(v_j)) - (1 - F_i(v_i^*(v_j))) \right] \frac{\partial v_i^*(v_j)}{\partial v_j}$
- $g(v_j) = \int_{v=v_j^*(1)}^{v_j} \theta_1(v) dv$
- $\theta_2(v_j) = \frac{g(v_j) f_j(v_j)}{(1-F_i(v_i^*(v_j)))(1-F_j(v_j))^2 \partial v_i^*(v_j)/\partial v_j}$

First, we show that there actually exists a transfer function that is incentive compatible if we require that the transfers are 0 if the good is not provided.

Lemma 2.5. *There exists an incentive compatible transfer function in which payments are only made if the good is provided*

$$t_i(v_i, v_j) \quad = \quad \begin{cases} v_i^*(v_j) - \frac{g(v_j)}{(1-F_i(v_i^*(v_j)))(1-F_j(v_j))} - \int_{v=v_i^*(v_j)}^{v_i} \theta_2(v_j^*(v)) dv & if\, p = 1 \\ 0 & otherwise \end{cases}$$

and

$$t_j(v_i, v_j) = \begin{cases} c - t_i(v_i, v_j) & if\, p = 1 \\ 0 & otherwise. \end{cases}$$

Note that this transfer function fulfills ex post budget balance. To find this transfer function we assume that $t_i(v_i, v_j)$ is additively separable, $t_i(v_i, v_j) = \phi(v_i) + \gamma(v_j)$, and then we derive a solution such that this function integrates to the expected transfers given in (2.6) and (2.7). The unique solution for the additive form is $\phi(v_i) = \int_{t=v_i}^{\overline{v}_i} \theta_2(v_j^*(t)) dt$ and $\gamma(v_j) = -\int_{t=v_i^*(v_j)}^{\overline{v}_i} \theta_2(v_j^*(t)) dt + v_i^*(v_j) - \frac{g(v_j)}{(1-F_i(v_i^*(v_j)))(1-F_j(v_j))}$. This leads to the transfer function given in Lemma 2.5. The details on how to calculate $t_i(v_i, v_j)$ are provided in the Appendix.

We have shown that there exists a transfer function such that no player gets a negative utility if the good is not provided. Next, we determine conditions under which this function also satisfies the property that no player pays more than her valuation if the good is provided. If both hold, there exists a transfer function such that the allocation rule is PostIR.

Lemma 2.6. *If the derivatives are* $\frac{\partial t_i(v_i,v_j)}{\partial v_i} \geq 0$ *and* $\frac{\partial t_i(v_i,v_j)}{\partial v_j} \leq 0$*, no player pays more than her valuation if the good is provided.*

Using that $t_i(v_i, v_j)$ is decreasing in v_j we know that $t_i(v_i, v_j)$ is maximized at $v_j^*(v_i)$ for a given v_i. Then it is sufficient to prove that $t_i(v_i, v_j^*) \leq v_i$. A similar argument applies for $t_j(v_i, v_j)$.

As a last step we show a sufficient condition on the distribution $F_i(v_i)$ such that the transfer function is monotone in both players' types.

Lemma 2.7. *If both hazard rates are monotone,* $\frac{\partial t_i(v_i,v_j)}{\partial v_i} \geq 0$ *and* $\frac{\partial t_i(v_i,v_j)}{\partial v_j} \leq 0$.

Note that the transfer function $t_i(v_i, v_j)$ also satisfies condition MON if both hazard rates are monotone. This implies Proposition 2.7.

2.5 Conclusion

This paper addresses a classical problem in Economics; the provision of public goods. Even though without commitment no first-best solution can be achieved, we show that there exists a simple game that can implement the same allocation rules as any mechanism in which players voluntary participate at all stages. This implies that a designer does not need to resort to complicated games to implement an outcome because the voluntary contribution game with communication can do the same job. Although players would like to free ride on the contribution of others, they have an incentive to reveal information about their own valuation to ensure that the good is provided at all.

Furthermore, we show that in a voluntary contribution game cheap talk can implement the same allocation rules as a mediator. For example, when several states negotiate to create a common financial fund to bail out banks, this constitutes a situation in which players have to decide on the provision of a public good and such a negotiation might resemble a voluntary contribution game with communication. Our analysis implies that having a mediator in such negotiations yields no advantage over direct communication.

Unfortunately, the class of Veto-IC mechanisms has not yet received sufficient attention in Economic Theory. Let us stress again that an ex post individual rationality constraint implies that a player has no regret ex post. To illustrate situations in which players voluntary participate in a mechanism at all stages and where they have the possibility to drop out, mechanism design should analyze Veto-IC mechanisms. Although ex post individually rational and Veto-IC mechanisms are often cumbersome, in some settings they might be crucial to find intuitive indirect mechanisms and to apply mechanism design to the real world.

2.6 Appendix

Proof of Proposition 2.1

Proof. Let r be an equilibrium outcome of the communication game with mediated communication.

(i) We show that r satisfies PostIR. Take two bids $b_i' = v_i$ and $b_i'' > v_i$. If $E[z(b_i'', b_{-i})] > E[z(b_i', b_{-i})]$, then for some b_{-i}, $b_i' + \sum b_{-i} = v_i + \sum b_{-i} < c$ and $b_i'' + \sum b_{-i} \geq c$ with $c = x_i(b_i'', b_{-i}) + \sum x_{-i}(b_i'', b_{-i})$. Combining this equality with the first inequality leads to $x_i(b_i'', b_{-i}) - v_i > \sum b_{-i} - \sum x_{-i}(b_i'', b_{-i})$. By the rules of the game no player has to pay more than her bid, $x_i(b) \leq b_i$ for all i, and it follows that $x_i(b_i'', b_{-i}) - v_i > \sum b_{-i} - \sum x_{-i}(b_i'', b_{-i}) \geq 0$ and therefore, $x_i(b_i'', b_{-i}) > v_i$. Thus, the player strictly prefers b_i' to b_i''.

If (a) $E[z(b_i', b_{-i})] = E[z(b_i'', b_{-i})] = 0$ or if (b) $E[z(b_i', b_{-i})] = E[z(b_i'', b_{-i})]$ and $E[x_i(b_i', b_{-i})] = E[x_i(b_i'', b_{-i})]$, the player is indifferent between b_i' and b_i''. In (a) the public good is never provided and $x_i = 0$ for all b_{-i}. Given the cost sharing rule of the game $x_i(b_i'', b_{-i}) \geq x_i(b_i', b_{-i})$ and no one pays more than her bid. Therefore in (b) it holds that $x_i(b_i'', b_{-i}) = x_i(b_i', b_{-i}) \leq b_i' < b_i''$. Thus, in equilibrium no player pays more than her valuation, $x_i \leq v_i$ for all v_{-i}, and r is PostIR.

(ii) The game satisfies ExpostBB by construction; $\sum x_i = c$ if the good is provided and $\sum x_i = 0$ otherwise.

(iii) We show that r satisfies Veto-IC. Suppose v_i sends $m_i \neq v_i$. Denote $\hat{b}_i$ the bid the mediator recommends to player i. Let $\tilde{b}_{-i}(m_i, \hat{b}_i)$ be the distribution over the mediator's recommendations to the other players given m_i and $\hat{b}_i$. Let $\tilde{b}_i(m_i)$ be the distribution over a player's own recommended bid given her message m_i. Let $\hat{u}_i(v_i, b_i, m_i, \hat{b}_i) = E[u_i(v_i, b_i, \tilde{b}_{-i}(m_i, \hat{b}_i))]$ be the expected utility of player i if she bids b_i given her message m_i and the recommendation $\hat{b}_i$.

Since the allocation rule r is an outcome of the game with a mediator, it has to hold that a player cannot be better off announcing a different type to the mediator which implies that

$$E\left[\sup_{b_i} \hat{u}_i\left(v_i, b_i, m_i, \tilde{b}_i\left(m_i\right)\right)\right] \leq U_i\left(v_i, v_i\right). \tag{2.8}$$

After any announcement and recommended action the player can choose how much to bid. Denote the optimal bid with β. Assume player i chooses $b_i' = \min\{\hat{b}_i, v_i\}$. This means he follows the recommended bid as long as it is not larger than his valuation. This gives us

$$\begin{aligned}
\hat{u}_i\left(v_i, \beta\left(v_i, m_i, \hat{b}_i\right), m_i, \hat{b}_i\right) &\geq \hat{u}_i\left(v_i, b_i', m_i, \hat{b}_i\right) \\
&= E\left[u_i\left(v_i, b_i', \tilde{b}_{-i}\left(m_i, \hat{b}_i\right)\right)\right] \\
&\geq E\left[\max\left\{0, u_i\left(v_i, \hat{b}_i, \tilde{b}_{-i}\left(m_i, \hat{b}_i\right)\right)\right\}\right]
\end{aligned}$$

where the second inequality follows from $u_i(v_i, \min\{v_i, \hat{b}_i(m_i, v_{-i})\}, b_{-i}) \geq \max\{0, u_i(v_i, \hat{b}_i, b_{-i})\}$ for all possible b_{-i}. If player i bids b_i' , inequality (2.8) leads to

$$E\left[E\left[\max\left\{0, u_i\left(v_i, \hat{b}_i, \tilde{b}_{-i}\left(m_i, \hat{b}_i\right)\right)\right\} \mid \tilde{b}_i\left(m_i\right)\right]\right] \leq U_i\left(v_i, v_i\right).$$

Let $\tilde{B}_{-i}(m_i)$ be the distribution over the other players' recommendations given player i sends message m_i.

$$\begin{aligned}
U_i\left(v_i, v_i\right) &\geq E\left[E\left[\max\left\{0, u_i\left(v_i, \hat{b}_i, \tilde{b}_{-i}\left(m_i, \hat{b}_i\right)\right)\right\} \mid \tilde{b}_i\left(m_i\right)\right]\right] \\
&= E\left[\max\left\{0, u_i\left(v_i, \tilde{b}_i\left(m_i\right), \tilde{B}_{-i}\left(m_i\right)\right)\right\}\right] && (2.9) \\
&= E\left[\max\left\{0, \pi_i\left(v_i, \tilde{Z}\left(m_i\right), \tilde{X}\left(m_i\right)\right)\right\}\right] && (2.10) \\
&= U_i^*\left(m_i, v_i\right)
\end{aligned}$$

In line 2.10 we write the player's utility π_i in terms of the allocation rule if she sends message m_i. This is equivalent to the players utility u_i in the game (line 2.9) because the player follows the recommended bid after sending message m_i. By PostIR $U_i^*(v_i, v_i) = U_i(v_i, v_i)$ and thus, r satisfies Veto-IC. □

Proof of Proposition 2.2

Proof. First, we show how to create a jointly controlled lottery and afterwards we prove the proposition.

At the communication stage every player sends one y_i. Take two announcements y_i and y_j and create a new variable $\tilde{y} \in [0, 1]$, where

$$\tilde{y} = \begin{cases} y_i + y_j & if\ y_i + y_j \leq 1 \\ y_i + y_j - 1 & if\ y_i + y_j > 1 \end{cases}$$

The new variable $\tilde{y}$ is uniformly distributed on $[0,1]$ if at least one of the variables y_i and y_j is uniformly distributed on $[0,1]$. Assume that y_i is uniformly distributed on $[0,1]$ and that $y_j \in [0,1]$. For $x \in [0,1]$

$$\begin{aligned} Pr\left[\tilde{y} \leq x \mid y_j\right] &= Pr\left[(y_i + y_j \leq x) \cup (1 < y_i + y_j \leq 1 + x) \mid y_j\right] \\ &= Pr\left[(y_i \leq x - y_j) \cup (1 - y_j < y_i \leq 1 + x - y_j) \mid y_j\right] \\ &= x \end{aligned}$$

independent of y_j. This implies that $Pr[\tilde{y} \leq x] = x$ and that $Pr[\tilde{y} \leq x \mid y_j \leq z] = x$.

Using the inverse transformation method for a one dimensional variable we can create a new random variable with any distribution out of a uniformly distributed random variable. The inverse transformation method works as follows: If a random variable x is distributed according to some F and y is uniformly distributed on $[0,1]$, then the random variable $F^{-1}(y)$ is also distributed according to F. Note that the inverse transformation method is defined for one dimensional random variables.

Since each player announces one y_i, we can create several random variables which are all uniformly distributed on $[0,1]$ by always combining announcement y_i and y_{i+1} to get $\tilde{y}_i$. Next, we need to show that these $\tilde{y}_i$ are independent. Take, e.g., $\tilde{y}_1$ which is constructed by y_1 and y_2 and $\tilde{y}_2$ which is constructed by y_2 and y_3. Assume that y_1 and y_3 are uniformly distributed on $[0,1]$ and $y_2 \in [0,1]$.

$$\begin{aligned} Pr\left[(\tilde{y}_1 \leq x) \cap (\tilde{y}_2 \leq z) \mid y_2\right] &= Pr\left[\tilde{y}_1 \leq x | \tilde{y}_2 \leq z, y_2\right] Pr\left[\tilde{y}_2 \leq z \mid y_2\right] \\ &= Pr\left[(y_1 \leq x - y_2) \cup (1 - y_2 < y_1 \leq 1 + x - y_2) \mid \tilde{y}_2 \leq z, y_2\right] \\ &\quad\; Pr\left[\tilde{y}_2 \leq z \mid y_2\right] \\ &= x Pr\left[\tilde{y}_2 \leq z \mid y_2\right] \\ &= Pr\left[\tilde{y}_1 \leq x \mid y_2\right] Pr\left[\tilde{y}_2 \leq z \mid y_2\right]. \end{aligned}$$

Now we show that we can construct a new $N-1$ dimensional variable with any distribution $G(a_1, a_2, ..., a_{N-1})$ using $N-1$ uniformly distributed variables. Take the marginal distribution of a_1, $G_{a_1}(a_1)$, and use one of the random variables $\tilde{y}_1$ to create with the inverse transformation method a new random variable with distribution $G_{a_1}(a_1)$. Given the realization of this first random variable we have determined a_1. Given a_1, there is a conditional distribution

$G_{a|a_1}(a_2, ..., a_{N-1}|a_1)$ which we denote by $\hat{G}(a_2, ..., a_{N-1})$. With a second random variable $\tilde{y}_2$ we can create another random variable which has the same distribution as the marginal distribution of a_2, $\hat{G}_{a_2}(a_2)$. Given the realization of a_1 and a_2 we can determine the conditional distribution and the marginal distribution of a_3 and so on. We always take one $\tilde{y}_i$ and with the inverse transformation method we create a random variable with the same distribution as the marginal distribution of one dimension of the multidimensional random variable. In this way we can create a $N-1$ dimensional variable with any distribution.

Take any allocation rule r that is Veto-IC, PostIR and ExpostBB. Now we construct an equilibrium of the unmediated communication game with outcome r. We show hat there exist an equilibrium in which (i) all players reveal their true type, (ii) send a y_i which is uniformly distributed and (iii) the players implement the same allocations as in r by bidding $b_i = x_i$ in the bidding stage. To prove that this in indeed an equilibrium we check that no player has an incentive to deviate.

(i) We have defined an allocation rule r as a mapping $V^N \to \Delta(\{0,1\} \times \mathbb{R}^N)$. The allocation rule r is PostIR and therefore we can restrict the transfers to $x_i \in [-Nc, c]$ if the good is provided and to $x_i = 0$ otherwise. We know that r is ex post budget balanced, which means that the transfers sum up to costs c if the good is provided and to 0 otherwise. Thus, if we know the transfers of $N-1$ players, the realization of $z \in \{0,1\}$ and the transfer of player N are pinned down as well. We haven shown before that we can construct a random variable with any distribution in $\mathbb{R}^{N-1}$ by $N-1$ uniformly distributed variables. If r is non degenerated, the realization of the random variable determines which allocation is implemented. The allocation is implemented by every player bidding $b_i = x_i$. Since r always implements an equilibrium of the stage game under complete information, no player can profit from deviation after everybody revealed her true type.

(ii) During the communication phase, no player has an incentive to change his announcement y_i, because the distribution of every $\tilde{y}_i$ being uniformly distributed on $[0,1]$ cannot be influenced by unilateral deviation and thus, the distribution of the jointly controlled lottery can neither be influenced by unilateral deviation. The distribution of the jointly controlled lottery is chosen such that for every v each allocation has the same probability in the lottery as in the allocation rule r.

(iii) It is left to show that players are willing to reveal their true type, $m_i = v_i$. After announcing $m_i \neq v_i$, the best player i can do is to bid 0 for all m_i, v_{-i} and outcomes of the jointly controlled lottery for which no public good is provided in equilibrium. For all cases in which the public good is provided in equilibrium, the player bids either as being type m_i or less. Since in the equilibrium of the stage game with complete information the bids exactly add up to the costs, $\sum b_i = c$, the good is not provided as soon as player i bids less than a player of type m_i would bid. Thus, her expected utility of announcing m_i is $E[\max\{0, v_i z(\tilde{\beta}_i(m_i), \tilde{\beta}_{-i}(m_i)) - x_i(\tilde{\beta}_i(m_i), \tilde{\beta}_{-i}(m_i))\}]$ where $\tilde{\beta}_i(m_i)$ is the distribution of a player's bid if she is type m_i and $\tilde{\beta}_{-i}(m_i)$ is the distribution of the other players' bids if player i announces m_i. Since the allocation rule r is Veto-IC and Exp IR, it is incentive compatible for a player to reveal her type truthfully to a mechanism given the allocation rule r and given that she can drop out and ensure a utility of zero. This implies that $E[\max\{0, \pi_i(v_i, \tilde{Z}(m_i), \tilde{X}(m_i))\}] \leq U(v_i, v_i)$. Therefore, it is also incentive compatible for every player to reveal her type truthfully at the cheap talk stage before the voluntary contribution game. $\square$

Example 2

The payoffs in the game are such that player i is not willing to reveal being type B using cheap talk because the other player j will use this information against i if j is type A. This loss is larger for i than the possible gain from revealing to be type B if j is also type B. The probability of being type A is larger than one half.

A,A

	X	Y
X	1,1	0,0
Y	0,0	0,0

B,B

	X	Y
X	0,0	0,0
Y	0,0	1,1

A,B

	X	Y
X	0,0	-2,0
Y	1,-2	-2,-3

B,A

	X	Y
X	0,0	-2,1
Y	0,-2	-3,-2

Proof of Proposition 2.4

Proof. Define $p^*(m_i, v_i)$ as the probability that the good is provided if player i sends message m_i and can reject the allocation. Define analogously $t_i^*(m_i, v_i)$ as the expected transfer. To simplify notation define $p(m_i, v_{-i}) \equiv p(m)$, $t_i(m_i, v_{-i}) \equiv t_i(m)$ and drop the index i for v_i and m_i. Define a function

$$D(\hat{v}, v', m) \equiv p^*(m, \hat{v})v' - t_i^*(m, \hat{v}) - U(v', v').$$

$D(\hat{v}, v', m)$ determines how much a player of type v' gains from sending message m and behaving as if she were $\hat{v}$. We are interested in $D(\hat{v}, v', m)$ for values v' in the interval $[\hat{v}, m]$.

Lemma 2.8. *If $D(\hat{v}, \hat{v}, m) > 0$, $D(\hat{v}, v', m) > D(\hat{v}, \hat{v}, m)$ for all $v' > \hat{v}$.*

Proof. First, we show that $D(\hat{v}, \hat{v}, m) > 0$ implies $p^*(m, \hat{v}) > p(\hat{v})$ and afterwards we show that this again yields to $D(\hat{v}, v', m) > D(\hat{v}, \hat{v}, m)$ for all $v' > \hat{v}$.

If $D(\hat{v}, \hat{v}, m) > 0$, it holds that $p^*(m, \hat{v})\hat{v} - t_i^*(m, \hat{v}) - (p(\hat{v})\hat{v} - t_i(\hat{v})) > 0$, which can be written as $(p^*(m, \hat{v}) - p(\hat{v}))\hat{v} - t_i^*(m, \hat{v}) + t_i(\hat{v}) > 0$. Define the sets $S_{\hat{v}} = \{v_{-i} | p(\hat{v}, v_{-i}) = 1\}$ and $S_m = \{v_{-i} | p(m, v_{-i}) = 1\}$ which are the sets of v_{-i} where the good is provided if a player announces type $\hat{v}$ respectively m and cannot reject the allocation. By MON (iii) the transfers for player i after announcing m are higher in $S_m \setminus S_{\hat{v}}$ than in $S_{\hat{v}}$ if announces being $\hat{v}$. Further, also for every v_{-i} in $S_{\hat{v}}$ a player has to pay weakly more if she announces a higher type. The proof is by contradiction. Assume $p^*(m, \hat{v}) - p(\hat{v}) \leq 0$. Then it cannot be strictly better for an agent to announce m if she is type $\hat{v}$ (it cannot be that $D(\hat{v}, \hat{v}, m) > 0$), because the probability that the good is provided is weakly lower and the expected transfer is weakly higher than if the player announces her true type $\hat{v}$.

If $p^*(m, \hat{v}) > p(\hat{v})$, it follows that $D(\hat{v}, v', m) > D(\hat{v}, \hat{v}, m)$ for $v' > \hat{v}$. To see this, rewrite $D(\hat{v}, v', m) > D(\hat{v}, \hat{v}, m)$ as

$$\begin{aligned}
p^*(m, \hat{v})v' - t_i^*(m, \hat{v}) - U(v', v') &> p^*(m, \hat{v})\hat{v} - t_i^*(m, \hat{v}) - U(\hat{v}, \hat{v}) \\
p^*(m, \hat{v})(v' - \hat{v}) &> p(v')v' - t_i(v') - (p(\hat{v})\hat{v} - t_i(\hat{v})) \\
t_i(v') - t_i(\hat{v}) &> p(v')v' - p(\hat{v})\hat{v} - p^*(m, \hat{v})(v' - \hat{v}) \qquad (2.11)
\end{aligned}$$

From incentive compatibility $v'p(v') - t_i(v') > v'p(\hat{v}) - t_i(\hat{v})$ it follows that $v'(p(v') - p(\hat{v})) > t_i(v') - t_i(\hat{v})$ which together with inequality (2.11) yields

$$\begin{aligned}
v'(p(v') - p(\hat{v})) &> p(v')v' - p(\hat{v})\hat{v} - p^*(m, \hat{v})(v' - \hat{v}) \\
p^*(m, \hat{v})(v' - \hat{v}) &> p(\hat{v})(v' - \hat{v}).
\end{aligned}$$

This inequality is true because $p^*(m, \hat{v}) > p(\hat{v})$ and $v' > \hat{v}$. $\square$

With this Lemma, we can prove the proposition. IC implies that $U(v, v) \geq U(m, v)$. Assume player i sends $m < v$. Since r is PostIR, $p(m)m - t_i(m) \geq 0$ and this implies that $p(m)v - t_i(m) \geq 0$ for all v_{-i}. Thus, $max\{0, p(m)v - t_i(m)\} = p(m)v - t_i(m)$ and Veto-IC follows from IC.

Now we have to show that player i does not send $m > v$. Fix a message m. If $p(m)v - t_i(m) \geq 0$ for all v_{-i}, player i is never going to opt out. IC implies that player i prefers to announce his true type to sending message m. However, if $p(m)v - t_i(m) < 0$ for some v_{-i} the player prefers to reject some allocations and to ensure 0. Using the notation defined before we can write

$$\begin{aligned} U^*(m, v) &= E_{v_{-i}}[max\{0, vp(m) - t_i(m)\}] \\ &= p^*(m, v)v - t_i^*(m, v). \end{aligned}$$

Note that $\forall v' \in [\hat{v}, m] : D(v', v', m) \geq D(\hat{v}, v', m)$. Suppose there $\exists \hat{v}$, $\hat{v} < m$, such that $D(\hat{v}, \hat{v}, m) > 0$. By Lemma 2.8 it follows that $D(m, m, m) \geq D(\hat{v}, m, m) > D(\hat{v}, \hat{v}, m) > 0$. However, since the allocation rule is PostIR, $D(m, m, m) = p^*(m, m)m - t_i^*(m, m) - U(m, m) = 0$ and this is a contradiction to $D(v, v, m) > 0$ at $v = \hat{v}$.

$\square$

Proof of Lemma 2.1

Proof. Take two allocation rules $\hat{r}$ and $\tilde{r}$ which both satisfy Veto-IC, ExpostBB and PostIR. We show that an allocation rule r, $r(v) = \lambda\hat{r}(v) + (1 - \lambda)\tilde{r}(v)$, with $\lambda \in [0, 1]$, also satisfies PostIR, ExpostBB and Veto-IC. Under $\hat{r}$ and $\tilde{r}$, $v_i z - x \geq 0$ for all v and all allocations which are implemented with positive probability. It follows directly that also under the allocation rule r the utility $v_i z - x \geq 0$ for all v and all allocations with positive probability. Thus, r is PostIR. Since $\hat{r}$ and $\tilde{r}$ are ExpostBB, also the convex combination of both is ExpostBB. It is left to show that r satisfies Veto-IC.

$$\begin{aligned}
U^*(v_i, v_i) =& E_r[max\{0, v_i z - x_i\}] \\
=& E_r[v_i z - x_i] \\
=& \lambda E_{\hat{r}}[v_i z - x_i] + (1-\lambda)E_{\tilde{r}}[v_i z - x_i] \\
=& \lambda E_{\hat{r}}[max\{0, v_i z - x_i\}] + (1-\lambda)E_{\tilde{r}}[max\{0, v_i z - x_i\}] \\
\geq& \lambda E_{\hat{r}}[max\{0, v_i z - x_i\}|m_i, v_i] + (1-\lambda)E_{\tilde{r}}[max\{0, v_i z - x_i\}|m_i, v_i] \\
\geq& E_r[max\{0, v_i z - x_i\}|m_i, v_i]
\end{aligned}$$

From line one to two and from line three to four we use that r, $\hat{r}$ and $\tilde{r}$ are PostIR. For the inequality in line five we use that $\hat{r}$ and $\tilde{r}$ satisfy Veto-IC. Since $U^*(v_i, v_i) \geq E_r[max\{0, v_i z - x_i\}|m_i, v_i]$, the allocation rule r fulfills Veto-IC. □

Proof of Proposition 2.5

Proof. We can write that a mechanism is incentive compatible iff for every $v_i \in [0, \overline{v}_i]$

$$U_i(v_i) \geq v_i \overline{p}_i(\hat{v}_i) - \overline{t}_i(\hat{v}_i)$$

for all $\hat{v}_i \in [0, \overline{v}_i]$. This means that no player has an incentive to lie about his true type. The mechanism is interim individual rational iff $U_i(v_i) \geq 0$ for every $v_i \in [0, \overline{v}_i]$. This ensures that every player voluntary participates in the mechanism. Note that every mechanism that is ex post individual rational is also interim individual rational, but not necessary the other way round.

From the incentive compatibility constrain for v_i and $\hat{v}_i$ we get

$$v_i \overline{p}_i(v_i) - \overline{t}_i(v_i) \geq v_i \overline{p}_i(\hat{v}_i) - \overline{t}_i(\hat{v}_i)$$

$$\hat{v}_i \overline{p}_i(\hat{v}_i) - \overline{t}_i(\hat{v}_i) \geq \hat{v}_i \overline{p}_i(v_i) - \overline{t}_i(v_i).$$

Combining these two inequalities leads to

$$(v_i - \hat{v}_i)\overline{p}_i(\hat{v}_i) \leq U_i(v_i) - U_i(\hat{v}_i) \leq (v_i - \hat{v}_i)\overline{p_i}(v_i). \tag{2.12}$$

From this inequality it follows that: if $v > \hat{v}_i$, it must hold that $\overline{p}_i(\hat{v}_i) \leq \overline{p}_i(v_i)$ and this implies that $\overline{p}_i(v_i)$ is increasing. Inequality (2.12) implies furthermore that $U_i'(v_i) = \overline{p}_i(v_i)$ at almost every v_i and $U_i(v_i) = U_i(0) + \int_0^{v_i} \overline{p}_i(s)ds$. It follows that $U_i(v_i)$ is increasing in v_i.

The following equalities hold

$$
\begin{aligned}
&\int (p(v) \sum v_i - \sum t_i(v)) f(v) dv \\
=&\sum \int (p(v)v_i - t_i(v)) f(v) dv \\
=&\sum \int_0^{\overline{v}_i} U_i(v_i) f_i(v_i) dv_i \\
=&\sum U_i(0) + \sum \left[\int_0^{\overline{v}_i} \int_0^{v_i} \overline{p}_i(s) ds f_i(v_i) dv_i \right] \\
=&\sum U_i(0) + \sum \left[\int_0^{\overline{v}_i} \int_s^{\overline{v}_i} \overline{p}_i(s) f_i(v_i) dv_i ds \right] \\
=&\sum U_i(0) + \sum \left[\int_0^{\overline{v}_i} \overline{p}_i(s)(1 - F_i(s)) ds \right] \\
=&\sum U_i(0) + \sum \left[\int_0^{\overline{v}_i} \int p(s, v_{-i}) f_{-i}(v_{-i}) dv_{-i} (1 - F_i(s)) ds \right] \\
=&\sum U_i(0) + \sum \left[\int_0^{\overline{v}_i} \int p(s, v_{-i}) f_{-i}(v_{-i}) f_i(s) \frac{1 - F_i(s)}{f_i(s)} dv_{-i} ds \right] \\
=&\sum U_i(0) + \int \left[\sum \frac{1 - F_i(v_i)}{f_i(v_i)} \right] f(v) p(v) dv.
\end{aligned}
$$

In the last step we replace s with v_i. Setting the first and the last row equal gives

$$
\begin{aligned}
&\sum U_i(0) \\
=&\int \left[p(v) \sum \left(v_i - \frac{1 - F_i(v_i)}{f_i} \right) - \sum t_i(v) \right] f(v) dv.
\end{aligned}
$$

Remember that this equality comes from the condition for incentive compatibility. For a mechanism to be interim individual rational the expected utility needs to be $U_i(v_i) \geq 0$, which leads to the necessary condition that $\sum U_i(0) \geq 0$. Only if (2.3) and (2.4) hold, the mechanism is incentive compatible and individual rational (necessary condition, but not sufficient).

The next step is to show that these conditions are sufficient. Suppose the provision rule $p(v)$ satisfies these conditions. We must find a payment function $t(v)$ such that the mechanism is incentive compatible and individual rational. Instead of searching for a transfer function that is interim IR and ex post BB we use a result by Börgers and Norman (2009). For independent types they prove that for every ex ante budget balanced mechanism there exists an ex post budget balanced mechanism with the same allocation rule and the same interim expected payments. Therefore, it is sufficient to find a transfer function that satisfies ex ante budget balance and interim IR. By incentive compatibility and the provision rule $p(v)$, the expected transfer $\bar{t}_i(v_i)$ is pinned down up to a constant $U_i(0)$. This is true for every player i. Pick $\sum U_i(0)$ such that ex ante budget balance is satisfied, $c\int p(v)f(v)dv = \sum\int \bar{t}_i(v_i)dv_i = \sum\int \left(v_i\bar{p}_i(v_i) - \int_0^{v_i}\bar{p}_i(s)ds\right)dv_i - \sum U_i(0)$. If $\sum U_i(0) > 0$, there are several solutions for $U_i(0)$. This determines the expected transfers $\bar{t}_i(v_i)$ and one possible transfer function is that every player always pays her expected transfer independent of the valuations of the other players, $t_i(v_i, v_{-i}) = \bar{t}_i(v_i)$.

This proves the proposition. □

Proof of Lemma 2.2

Proof. Assume that a mechanism $(p, \Gamma(p))$ is an efficient allocation rule. We maximize

$$\sum \mu_i \int_0^{\overline{v}_i} U_i(v_i)f_i(v_i)dv_i = \sum \mu_i U_i(0) + \sum\left[\mu_i \int_0^{\overline{v}_i}\int_0^{v_i}\overline{p}_i(s)dsf_i(v_i)dv_i\right]$$

and we know that $\Gamma(p) = \sum U_i(0)$. Thus the expression is maximized, if $U_{i,\mu^{max}}(0) = \Gamma(p)$ for the player with the highest welfare weight μ^{max}, and $U_i(0) = 0$ for all players with $\mu_i < \mu^{max}$. If several players have $\mu_i = \mu^{max}$, the sum of all these $U_{i,\mu^{max}}(0)$ has to be equal to $\Gamma(p)$. □

Proof of Proposition 2.6

Proof. First we define the provision function $p^\alpha(v)$

$$\begin{aligned} p(v) &= 1 \qquad && if \sum q_i(v_i, \alpha_i) \geq c \\ &= 0 && otherwise \end{aligned}$$

and $\alpha \equiv (\alpha_1, \ldots, \alpha_N)$. We choose $p(v)$ to maximize (2.5) under the constraint that $\Gamma(p) = \sum U_i(0) = \int\left(\sum q_i(v_i, 1) - c\right)p(v)f(v)dv \geq 0$. Analog to the proof of Proposition 2.5, (2.5) can be written as

$$\begin{aligned} &\sum \mu_i \int_0^{\overline{v}_i} U_i(v_i)f_i(v_i)dv_i \\ =&\sum \mu_i U_i(0) + \int\left[\sum \mu_i \frac{1 - F_i(v_i)}{f_i(v_i)}\right]f(v)p(v)dv. \end{aligned}$$

Using Lemma 2.2 we can write

$$
\begin{aligned}
=&\mu^{max}\Gamma(p)+\int\left[\sum \mu_i\frac{1-F_i(v_i)}{f_i(v_i)}\right]f(v)p(v)dv\\
=&\mu^{max}\int\left(\sum q_i(v_i,1)-c\right)p(v)f(v)dv+\int\left[\sum \mu_i\frac{1-F_i(v_i)}{f_i(v_i)}\right]f(v)p(v)dv\\
=&\int\left(\sum \mu^{max}q_i(v_i,1)+\sum \mu_i\frac{1-F_i(v_i)}{f_i(v_i)}-\mu^{max}c\right)p(v)f(v)dv\\
=&\mu^{max}\int\left[\sum q_i(v_i,\frac{\mu^{max}-\mu_i}{\mu^{max}})-c\right]p(v)f(v)dv.
\end{aligned}
$$

This expression is maximized for $\alpha_i=\frac{\mu^{max}-\mu_i}{\mu^{max}}$ and $p=1$ iff $\sum q_i(v_i,\alpha_i)\geq c$. If $\Gamma(p^\alpha)\geq 0$, this is the solution to the problem.

If $\Gamma(p^\alpha)<0$, the condition $\Gamma(p)\geq 0$ is binding and we use Lagrange

$$
\begin{aligned}
&\mu^{max}\Gamma(p)+\int\left[\sum \mu_i\frac{1-F_i(v_i)}{f_i(v_i)}\right]f(v)p(v)dv+\lambda\Gamma(p)\\
=&\int\left[(\mu^{max}+\lambda)\sum v_i-\sum(\mu^{max}+\lambda-\mu_i)\frac{1-F_i(v_i)}{f_i}-(\mu^{max}+\lambda)c\right]p(v)f(v)dv\\
=&(\mu^{max}+\lambda)\int\left[\sum\left(v_i-\frac{\mu^{max}+\lambda-\mu_i}{\mu^{max}+\lambda}\frac{1-F_i(v_i)}{f_i}\right)-c\right]p(v)f(v)dv.
\end{aligned}
$$

It follows that the Lagrangian is maximized by a provision rule p^α with $\alpha_i=\frac{\mu^{max}+\lambda-\mu_i}{\mu^{max}+\lambda}$ which is 1 if $\sum q_i(\alpha_i)-c\geq 0$ and 0 otherwise.

Now we have to show that there exists an λ such that $\Gamma(p^\alpha)=0$. By assumption $q_i(1)=v_i-\frac{1-F_i(v_i)}{f_i(v_i)}$ is increasing in v_i for all i. Then $q_i(\alpha_i)$ is also increasing in v_i for every α_i between 0 and 1. This implies that $p^\alpha(v)$ increases in v_i and in v_{-i}.[18] Note that every $\alpha_i=\frac{\mu^{max}+\lambda-\mu_i}{\mu^{max}+\lambda}$ is weakly increasing in λ. For $\lambda=0$, $\alpha_i=\frac{\mu^{max}-\mu_i}{\mu^{max}}$ and $\Gamma(p^\alpha)<0$ by assumption (if it is positive the argumentation of part one applies). For $\lambda\to\infty$, all $\alpha_i\to 1$ and p^1 is positive if $\sum q_i(1)\geq c$. It follows $\Gamma(p^1)\geq 0$. Remember that $\Gamma(p^\alpha)=\int\left(\sum q_i(1)-c\right)p^\alpha(v)f(v)dv$.

Now we show that $\Gamma(p^\alpha)$ is monotone and continuous in λ. $\sum q_i(\alpha_i)\geq c$ is decreasing in α_i, so p^α is also decreasing in every α_i. Take two values for α_i, α' and α'', with $\alpha'<\alpha''$, while keeping the values for α_{-i} constant. $\Gamma(p^{\alpha',\alpha_{-i}})$ differs from $\Gamma(p^{\alpha'',\alpha_{-i}})$ only because $0=p^{\alpha'',\alpha_{-i}}(v)<p^{\alpha',\alpha_{-i}}(v)=1$ for some v for which $q_i(\alpha'')+\sum\limits_{-i}q_i(\alpha)<c$ and hence $\sum q_i(1)<c$. Thus, for these valuations v, $\sum q_i(1)-c<0$ and because they are included in the integral of α' and not in the integral of α'' anymore, $\Gamma(p^{\alpha_i,\alpha_{-i}})$ is increasing in α_i. And since α_i is weakly increasing in λ,$\Gamma(p^\alpha)$ is increasing in λ.

[18]Note that $\overline{p}_i(v_i)$ is also increasing in v_i.

By assumption $q_i(1)$ is increasing in v_i and it follows that $q_i(\alpha)$ is strictly increasing in v_i. Therefore, if $\sum\limits_{-i} q_i(\alpha_i) < c$, $\sum q_i(\alpha_i) = c$ has at most one solution in v_i for a given v_{-i} and α. Denote this solutions with $v_i^*(v_{-i}, \alpha)$ and it is continuous in α and v_{-i}. Then we can write $\Gamma(p^\alpha) = \int \int_{v_i^*(v_{-i})}^{\overline{v}_i} \left(\sum q_i(v_i, 1) - c\right) p(v) f(v) dv_i dv_{-i}$ which is also continuous in α and therefore $\Gamma(p^\alpha)$ is continuous in λ.

It follows that for some λ it has to hold that $\Gamma(p^\alpha) = 0$. □

Proof of Lemma 2.3

Proof. Define the set S of v_{-i} as $S(\underline{w}, \overline{w}) = \left\{ v_{-i} \mid \underline{w} \leq \sum\limits_{-i} q_i(\alpha_i) \leq \overline{w} \right\}$. Split the integral in $\Gamma(p)$ in a first part in which the good is always provided independent of the valuation of player i, given the valuations v_{-i}, and a second part in which the provision depends on v_i. Define $v_i^*(v_{-i})$ in the same way as before (as the solution in v_i to $\sum q_i(\alpha_i) = c$) and w^* as $w^* = c - q_i(\overline{v}_i, \alpha_i)$, which is the minimum of $\sum\limits_{-i} q_i(\alpha_i)$ such that the good is provided with positive probability. Note that $q_i(\overline{v}_i, \alpha_i) = \overline{v}_i$.

$$
\begin{aligned}
\Gamma(p) &= \int \left(\sum q_i(v_i,1) - c\right) p(v) f(v) dv \\
&= \int_{S(c,\sum_{-i}\overline{v}_i)} \int_0^{\overline{v}_i} \left(\sum q_i(v_i,1) - c\right) f_i(v) f_{-i}(v_{-i}) dv_i dv_{-i} \\
&+ \int_{S(w^*,c)} \int_{v_i^*(v_{-i})}^{\overline{v}_i} \left(\sum q_i(v_i,1) - c\right) f_i(v) f_{-i}(v_{-i}) dv_i dv_{-i} \\
&= \int_{S(c,\sum_{-i}\overline{v}_i)} \int_0^{\overline{v}_i} \left(v_i - \frac{1-F_i(v_i)}{f_i(v_i)} + \sum_{-i} q_i(v_i,1) - c\right) f_i(v) f_{-i}(v_{-i}) dv_i dv_{-i} \\
&+ \int_{S(w^*,c)} \int_{v_i^*(v_{-i})}^{\overline{v}_i} \left(v_i - \frac{1-F_i(v_i)}{f_i(v_i)} + \sum_{-i} q_i(v_i,1) - c\right) f_i(v) f_{-i}(v_{-i}) dv_i dv_{-i} \\
&= \int_{S(c,\sum_{-i}\overline{v}_i)} \left[v_i(F_i(v_i) - 1) + (\sum_{-i} q_i(v_i,1) - c) F_i(v_i)\right]_0^{\overline{v}_i} f_{-i}(v_{-i}) dv_{-i} \\
&+ \int_{S(w^*,c)} \left[v_i(F_i(v_i) - 1) + (\sum_{-i} q_i(v_i,1) - c) F_i(v_i)\right]_{v_i^*(v_{-i})}^{\overline{v}_i} f_{-i}(v_{-i}) dv_{-i} \\
&= \int_{S(c,\sum_{-i}\overline{v}_i)} (\sum_{-i} q_i(v_i,1) - c) f_{-i}(v_{-i}) dv_{-i} \\
&+ \int_{S(w^*,c)} \left((1 - F_i(v_i^*(v_{-i})))(v_i^*(v_{-i}) + \sum_{-i} q_i(v_i,1) - c)\right) f_{-i}(v_{-i}) dv_{-i}.
\end{aligned}
$$

Remember that $\alpha_i = 0$ for the player with the highest welfare weight μ_i^{max} if $\Gamma(p^\alpha) \geq 0$. Assume that for all other players $\alpha_i = 1$ and denote this α as α^{max}. $\Gamma(p^{\alpha^{max}}) = \int_{S(c,\sum_{-i}\overline{v}_i)} (\sum_{-i} q_i(v_i,1) - c) f_{-i}(v_{-i}) dv_{-i} + 0$ and the second term cancels because $(v_i^*(v_{-i}) + \sum_{-i} q_i(v_i,1) - c) = 0$. Thus $\Gamma(p^{\alpha^{max}}) > 0$ if $\sum_{-i}\overline{v}_i > c$ and $\Gamma(p^{\alpha^{max}}) = 0$ otherwise. If $\sum_{-i}\overline{v}_i \leq c$ for all i, which implies that every player is crucial, then $\Gamma(p^{\alpha^{max}}) = 0$ independent which player has $\alpha_i = 0$ in α^{max}. Remember that $\Gamma(p^\alpha)$ is increasing in α_i. This implies that if we lower any α_i to be smaller than 1, $\Gamma(p^\alpha) < 0$. Therefore $\Gamma(p) = 0$ is binding for all welfare weights if every player is crucial.

To complete the proof note that $\alpha_i = 0$ for all i if all players have equal welfare weights. However, $\Gamma(p^0) < 0$[19] and therefore $\Gamma(p) = 0$ is binding. □

[19]Therefore, no ex post efficient allocation rule is possible under interim IR and budget balance.

Proof of Lemma 2.4

Proof. From incentive compatibility we know that

$$U_i(v_i) = U_i(0) + \int_0^{v_i} \overline{p}(v)dv.$$

Together with $U_i(v_i) = \overline{p}(v_i)v_i - \overline{t}(v_i)$ this gives the following condition for the expected transfers of player i

$$\begin{aligned}
\overline{t}_i(v_i) &= \int_{v=0}^{v_i} v d\overline{p}(v) \\
&= \overline{p}(v_i)v_i - \int_{v=0}^{v_i} \overline{p}(v)dv - U_i(0) \\
&= \overline{p}(v_i)v_i - \int_{v=0}^{v_i} \int_{v_j=0}^{\overline{v}_j} p(v, v_j) f_j(v_j) dv_j dv.
\end{aligned}$$

We have shown before that in an efficient allocation rule with two players $U_i(0) = 0$. Using that the allocation rule is deterministic and monotone, it follows that $p(v, v_j) = 1$ if $v_j \geq v_j^*(v)$.

$$\begin{aligned}
\overline{t}_i(v_i) &= \overline{p}(v_i)v_i - \int_{v=v_i^*(\overline{v}_j)}^{v_i} \int_{v_j=v_j^*(v)}^{\overline{v}_j} f_j(v_j)dv_j dv \\
&= \overline{p}(v_i)v_i - \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} \int_{v=v_i^*(v_j)}^{v_i} f_j(v_j)dv dv_j \\
&= \overline{p}(v_i)v_i - \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} [v f_j(v_j)]_{v_i^*(v_j)}^{v_i} dv_j \\
&= v_i \int_{v_j=0}^{\overline{v}_j} p(v, v_j) f_j(v_j)dv_j - \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i f_j(v_j)dv_j + \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j) f_j(v_j)dv_j \\
&= \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j) f_j(v_j)dv_j
\end{aligned}$$

□

Proof of Lemma 2.5

Proof. In order to show that this transfer function is incentive compatible, we show that $\int_{v_j=0}^{\overline{v}_j} t_i(v_i, v_j) f_j(v_j)dv_j = \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j) f_j(v_j)dv_j$ is true for the transfer function given in the proposition.

We take the derivative with respect to v_i of both sides of the equation. For the l.h.s we use that $\int_{v_j=0}^{\overline{v}_j} t_i(v_i,v_j)f_j(v_j)dv_j = \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} t_i(v_i,v_j)f_j(v_j)dv_j$ and get the derivative

$$\begin{aligned}
& \frac{\partial \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} t_i(v_i,v_j)f_j(v_j)dv_j}{\partial v_i} \\
= & \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} \frac{\partial t_i(v_i,v_j)}{\partial v_i} f_j(v_j)dv_j - t_i(v_i,v_j^*(v_i))f_j(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i} \\
= & \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} (-\theta_2(v_j^*(v_i)))f_j(v_j)dv_j - v_i f_j(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i} \\
& + \frac{g(v_j^*(v_i))f_j(v_j^*(v_i))}{(1-F_i(v_i))(1-F_j(v_j^*(v_i)))}\frac{\partial v_j^*(v_i)}{\partial v_i} \\
= & (1-F_j(v_j^*(v_i))(-\theta_2(v_j^*(v_i))) - v_i f_j(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i} \\
& + \frac{g(v_j^*(v_i))f_j(v_j^*(v_i))}{(1-F_i(v_i))(1-F_j(v_j^*(v_i)))}\frac{\partial v_j^*(v_i)}{\partial v_i} \\
= & -v_i f_j(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i}.
\end{aligned}$$

The derivative of the r.h.s. is

$$\frac{\partial \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j)f_j(v_j)dv_j}{\partial v_i} = -\frac{\partial v_j^*(v_i)}{\partial v_i} v_i f_j(v_j^*(v_i)).$$

The derivatives of both sides are the same. For v_i such that $v_j^*(v_i) = \overline{v}_j$ the value of both functions is the same. Thus, both expressions for the expected transfer are equivalent.

Next, we show that the transfer function is incentive compatible for player j

$$\int_{v_i=v_i^*(v_j)}^{\overline{v}_i} (c - t_i(v_i,v_j))f_i(v_i)dv_i = \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} v_j^*(v_i)f_i(v_i)dv_i. \tag{2.13}$$

To take the derivatives of both sides, first we calculate the derivative $\frac{\partial t_i(v_i,v_j)}{\partial v_j}$ separately

$$\begin{aligned}
& \frac{\partial t_i(v_i, v_j)}{\partial v_j} = \frac{\partial \gamma(v_j)}{\partial v_j} \\
= \; & \theta_2(v_j)\frac{\partial v_i^*(v_j)}{\partial v_j} + \frac{\partial v_i^*(v_j)}{\partial v_j} - \frac{(1 - F_i(v_i^*(v_j))(1 - F_j(v_j)\theta_1(v_j)}{(1 - F_i(v_i^*(v_j)))^2(1 - F_j(v_j))^2} \\
& + \frac{g(v_j)\left[(-f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}(1 - F_j(v_j)) + (1 - F_i(v_i^*(v_j))(-f_j(v_j))\right]}{(1 - F_i(v_i^*(v_j)))^2(1 - F_j(v_j))^2} \\
= \; & \frac{\partial v_i^*(v_j)}{\partial v_j} - \frac{\theta_1(v_j)}{(1 - F_i(v_i^*(v_j))(1 - F_j(v_j)} - \frac{g(v_j)f_i(v_i^*(v_j)\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j)))^2(1 - F_j(v_j)} \\
= \; & \frac{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j)} - \frac{g(v_j)f_i(v_i^*(v_j)\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j)))^2(1 - F_j(v_j)} \\
& + \frac{(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))\left[(v_j + v_i^*(v_j) - c)f_i(v_i^*(v_j)) - (1 - F_i(v_i^*(v_j)))\right]\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j)} \\
= \; & \frac{\left[-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))\right]f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j)}.
\end{aligned}$$

Now the derivative of the l.h.s of (2.13) is

$$\begin{aligned}
& \frac{\partial \int_{v_i=v_i^*(v_j)}^{\overline{v}_i}(c - t_i(v_i, v_j))f_i(v_i)dv_i}{\partial v_j} \\
= \; & \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} \frac{-\partial t_i(v_i, v_j)}{\partial v_j}f_i(v_i)dv_i - (c - t_i(v_i^*(v_j), v_j))\, f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j} \\
= \; & \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} \left(\frac{-\left[-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))\right]f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j)}\right) \\
& f_i(v_i)dv_i - f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}\left(c - v_i^* + \frac{g(v_j)}{(1 - F_j(v_j))(1 - F_i(v_i^*(v_j)))}\right) \\
= \; & \left(\frac{\left[g(v_j) - (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))\right]f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1 - F_i(v_i^*(v_j))^2(1 - F_j(v_j)}\right) \\
& (1 - F_i(v_i^*(v_j))) - f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}\left(c - v_i^* + \frac{g(v_j)}{(1 - F_j(v_j))(1 - F_i(v_i^*(v_j)))}\right) \\
= \; & -v_j f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}.
\end{aligned}$$

The derivative of the r.h.s of (2.13) is

$$\frac{\partial \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} v_j^*(v_i)f_i(v_i)dv_i}{\partial v_j} = -\frac{\partial v_i^*(v_j)}{\partial v_j}v_j f_i(v_i^*(v_j)).$$

Analogously to before, pick v_j such that $v_i^*(v_j) = \overline{v}_i$ and the value of l.h.s and the r.h.s of (2.13) is the same. It follows that the transfer functions in the proposition are incentive compatible. □

Idea how to find the transfer function in Proposition 2.5

1. Take the derivative of

$$\int_{v_j=v_j^*(v_i)}^{\overline{v}_j} (\phi(v_i) + \gamma(v_j)) f_j(v_j) dv_j = \int_{v_j=v_j^*(v_i)}^{\overline{v}_j} v_i^*(v_j) f_j(v_j) dv_j$$

with respect to v_i. Substitute $v_i = v_i^*(v_j)$ and solve for $\gamma(v_j)$.

$$\gamma(v_j^*(v_i)) = -\phi(v_i) + v_i + \phi'(v_i) \frac{1 - F_j(v_j^*(v_i))}{\frac{\partial v_j^*}{\partial v_i} f_j(v_j^*(v_i))}$$

$$\gamma(v_j) = -\phi(v_i^*(v_j)) + v_i^*(v_j) + \phi'(v_i^*(v_j)) \frac{1 - F_j(v_j)}{f_j(v_j)} \frac{\partial v_i^*}{\partial v_j}$$

2. Take the derivative of $\gamma(v_j)$ with respect to v_j.

$$\begin{aligned} \gamma'(v_j) &= \frac{\partial v_i^*}{\partial v_j} + \phi'(v_i^*(v_j)) \left[\frac{1 - F_j(v_j)}{f_j(v_j)} \frac{\partial^2 v_i^*}{\partial v_j^2} + \frac{\partial v_i^*}{\partial v_j} \left(-2 - \frac{f_j'(v_j)(1 - F_j(v_j))}{f_j(v_j)^2} \right) \right] \\ &\quad + \phi''(v_i^*(v_j)) \frac{1 - F_j(v_j)}{f_j(v_j)} \left(\frac{\partial v_i^*}{\partial v_j} \right)^2 \end{aligned}$$

3. Take the derivative of

$$\int_{v_i=v_i^*(v_j)}^{\overline{v}_i} (c - \phi(v_i) - \gamma(v_j)) f_i(v_i) dv_i = \int_{v_i=v_i^*(v_j)}^{\overline{v}_i} v_j^*(v_i) f_i(v_i) dv_i$$

with respect to v_j. Plug in $\gamma(v_j)$ and $\gamma'(v_j)$. This gives the differential equation

$$\begin{aligned}
&\phi'(v_i^*(v_j))\left[\frac{1-F_j(v_j)}{f_j(v_j)}f_i(v_i^*(v_j))\left(\frac{\partial v_i^*}{\partial v_j}\right)^2\right.\\
&\left.-(1-F_i(v_i^*(v_j)))\left(\frac{1-F_j(v_j)}{f_j(v_j)}\frac{\partial^2 v_i^*}{\partial v_j^2}+\frac{\partial v_i^*}{\partial v_j}\left(-2-\frac{f_j'(v_j)(1-F_j(v_j))}{f_j(v_j)^2}\right)\right)\right]\\
&-\phi''(v_i^*(v_j))(1-F_i(v_i^*(v_j)))\frac{1-F_j(v_j)}{f_j(v_j)}\left(\frac{\partial v_i^*}{\partial v_j}\right)^2\\
=&-\frac{\partial v_i^*}{\partial v_j}\left[f_i(v_i^*(v_j))(v_j+v_i^*(v_j)-c)-(1-F_i(v_i^*(v_j)))\right].
\end{aligned}$$

4. $\phi(v_i)=\int_{t=v_i}^{\overline{v}_i}\theta_2(v_j^*(t))dt$ is a solution. Plug $\phi(v_i^*(v_j))$ and $\phi'(v_i^*(v_j))$ into $\gamma(v_j)$, this leads to $\gamma(v_j)=-\int_{t=v_i^*(v_j)}^{\overline{v}_i}\theta_2(v_j^*(t))dt+v_i^*(v_j)-\frac{g(v_j)}{(1-F_i(v_i^*(v_j)))(1-F_j(v_j))}$.

Proof of Lemma 2.6

Proof. The derivative of $t_i(v_i,v_j)$ with respect to v_i is

$$\begin{aligned}
\frac{\partial t_i(v_i,v_j)}{\partial v_i} &= \frac{\partial\phi(v_i)}{\partial v_i}\\
&= -\theta_2(v_j^*(v_i))\\
&= -\frac{g(v_j^*(v_i))f_j(v_j^*(v_i))}{(1-F_i(v_i))(1-F_j(v_j^*(v_i)))^2}\frac{\partial v_j^*(v_i)}{\partial v_i}
\end{aligned}$$

and with respect to v_j we have derived in the proof of Proposition 2.5 that

$$\begin{aligned}
&\frac{\partial t_i(v_i,v_j)}{\partial v_j}=\frac{\partial\gamma(v_j)}{\partial v_j}\\
=&\frac{\left[-g(v_j)+(v_j+v_i^*(v_j)-c)(1-F_i(v_i^*(v_j))(1-F_j(v_j))\right]f_i(v_i^*(v_j))\frac{\partial v_i^*(v_j)}{\partial v_j}}{(1-F_i(v_i^*(v_j))^2(1-F_j(v_j)}.
\end{aligned}$$

Since $\frac{\partial v_j^*(v_i)}{\partial v_i}\leq 0$, the derivative of $t_i(v_i,v_j)$ w.r.t. v_i, $\frac{\partial t_i(v_i,v_j)}{\partial v_i}\geq 0$, is positive if and only if $g(v_j^*(v_i))\geq 0$. Furthermore, the derivative of $t_i(v_i,v_j)$ w.r.t. v_j is negative, $\frac{\partial t_i(v_i,v_j)}{\partial v_j}\leq 0$, if and only if $[-g(v_j)+(v_j+v_i^*(v_j)-c)(1-F_i(v_i^*(v_j))(1-F_j(v_j))\ ]\geq 0$. For a given v_i, the transfer t_i is decreasing in v_j. Thus, the transfer is maximized for $v_j=v_j^*(v_i)$ in which case it is $t_i(v_i,v_j^*(v_i))=v_i-\frac{g(v_j^*(v_i))}{(1-F_i(v_i))(1-F_j(v_j^*(v_i)))}\leq v_i$, because $g(v_j^*(v_i))\geq 0$. The transfer of player j, for a given valuation v_j, is maximized if $v_i=v_i^*(v_j)$. This is $t_j(v_i^*(v_j),v_j)=c-t_i(v_i^*(v_j),v_j)=c-v_i^*(v_j)+\frac{g(v_j)}{(1-F_i(v_i^*(v_j)))(1-F_j(v_j))}\leq v_j$ by $\frac{\partial t_i(v_i,v_j)}{\partial v_j}\leq 0$. □

Proof of Lemma 2.7

Proof. In the proof of Lemma 2.6 we use that (i) $\frac{\partial t_i(v_i,v_j)}{\partial v_i} \geq 0$ if and only if $g(v_j^*(v_i)) \geq 0$ and (ii) $\frac{\partial t_i(v_i,v_j)}{\partial v_j} \leq 0$ if and only if $-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j)) \geq 0$. In the following part, we show a sufficient condition on the distribution $F_i(v_i)$ such that (i) and (ii) hold. We derive the sufficient conditions for $g(v_j^*(v_i)) \geq 0$ and at the end we show that for $-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j)) \geq 0$ we can use a symmetric argument.

The proof has three steps: first, in Lemma 18, we show that $g(v_j^*(\overline{v}_i))$ and $g(\overline{v}_j)$ are weakly positive and second, in Lemma 19, we show that for intermediate v_i, $g(v_j^*(v_i))$ is also weakly positive if its derivative $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ changes its sign only once. Third, in Lemma 20, we show that $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ is first positive and then negative if the hazard rates are monotone. This proves that $g(v_j^*(v_i)) \geq 0$.

Lemma 2.9. $g(v_j^*(\overline{v}_i)) = 0$ *and* $g(\overline{v}_j) = \sum U_i(0) \geq 0$.

Proof. It follows from the definition of $g(v_j) = \int_{v=v_j^*(\overline{v}_i)}^{v_j} \theta_1(v)dv$ that $g(v_j^*(\overline{v}_i)) = 0$. For $\overline{v}_j$ we can write

$$\begin{aligned}
&g(\overline{v}_j) \\
&= \int_{v=v_j^*(\overline{v}_i)}^{\overline{v}_j} \theta_1(v)dv \\
&= \int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} \theta_1(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i}dv_i \\
&= -\int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} \left(1 - F_j\left(v_j^*(v_i)\right)\right)\left[\left(v_i + v_j^*(v_i) - c\right) f_i(v_i) - (1 - F_i(v_i))\right] dv_i. \qquad (2.14)
\end{aligned}$$

Compare this to the following expression that we have derived for incentive compatible mechanisms

$$
\begin{aligned}
&\sum U_i(0)\\
&= \int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} \int_{v_j^*(v_i)}^{\overline{v}_j} \left[(v_i + v_j - c) - \frac{1-F_i(v_i)}{f_i(v_i)} - \frac{1-F_j(v_j)}{f_j(v_j)} \right] f_i(v_i) f_j(v_j) dv_j dv_i\\
&= \int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} \int_{v_j^*(v_i)}^{\overline{v}_j} [(v_i - c) f_i(v_i) f_j(v_j) - (1 - F_i(v_i)) f_j(v_j)\\
&\quad + v_j f_i(v_i) f_j(v_j) - (1 - F_j(v_j)) f_i(v_i)] dv_j dv_i\\
&= \int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} [((v_i - c) f_i(v_i) - (1 - F_i(v_i))) F_j(v_j) - v_j (1 - F_j(v_j)) f_i(v_i)]_{v_j^*(v_i)}^{\overline{v}_j} dv_i\\
&= \int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} [((v_i - c) f_i(v_i) - (1 - F_i(v_i))) \left(1 - F_j\left(v_j^*(v_i)\right)\right)\\
&\quad + v_j^*(v_i)(1 - F_j\left(v_j^*(v_i)\right)) f_i(v_i)] dv_i.\\
&= -\int_{v=v_i^*(\overline{v}_j)}^{\overline{v}_i} \left(1 - F_j\left(v_j^*(v_i)\right)\right) \left[\left(v_i + v_j^*(v_i) - c\right) f_i(v_i) - (1 - F_i(v_i))\right] dv_i. \qquad (2.15)
\end{aligned}
$$

By individual rationality we know that this has to be positive. Expression (2.14) is equal to (2.15) and thus, $g(\overline{v}_j) \geq 0$. □

Lemma 2.10. *$g(v_j^*(v_i))$ is weakly positive in the interval from $\left[g(v_j^*(\overline{v}_i)), g(\overline{v}_j)\right]$ if $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ switches sign only once.*

Proof. Remember that $g(v_j^*(v_i))$ is weakly positive at both endpoints $v_i^*(\overline{v}_j)$ and $\overline{v}_i$. Furthermore, $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ is negative at $v_i = \overline{v}_i$ because $\frac{\partial g(v_j^*(v_i))}{\partial v_i} = \theta_1(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i}$ which is $-(1-F_j(v_j^*(\overline{v}_i)))\left[(\overline{v}_i + v_j^*(\overline{v}_i) - c) f_i(\overline{v}_i)\right]$. Thus, if $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ switches sign only once, $g(v_j^*(v_i))$ is positive in $\left[g(v_j^*(\overline{v}_i)), g(\overline{v}_j)\right]$. □

Lemma 2.11. *$\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ switches sign only once if the hazard rates are monotone and the allocation rule is efficient, e.g. has the form*

$$
\begin{aligned}
p(v) &= 1 \quad if \sum \left(v_i - \alpha_i \frac{1-F_i(v_i)}{f_i(v_i)} \right) \geq c\\
&= 0 \quad otherwise.
\end{aligned}
$$

Proof.

$$\begin{aligned}\frac{\partial g(v_j^*(v_i))}{\partial v_i} &= \theta_1(v_j^*(v_i))\frac{\partial v_j^*(v_i)}{\partial v_i}\\ &= (1-F_j(v_j^*(v_i)))\left[(v_i+v_j^*(v_i)-c)f_i(v_i)-(1-F_i(v_i))\right]\end{aligned}$$

It is equal to 0, if

$$\begin{aligned}(v_i+v_j^*(v_i)-c)f_i(v_i)-(1-F_i(v_i)) &= 0\\ \Leftrightarrow (v_i+v_j^*(v_i)-c)-\frac{1-F_i(v_i)}{f_i(v_i)} &= 0.\end{aligned}$$

If $v_i=\overline{v}_i$, the l.h.s is positive because $(\overline{v}_i+v_j^*(\overline{v}_i)-c)\geq 0$. The l.h.s is negative if $v_i=v_i^*(\overline{v}_j)$, because $\overline{v}_j-\alpha_j\frac{1-F_j(\overline{v}_j)}{f_j(\overline{v}_j)}+v_i^*(\overline{v}_j)-\alpha_i\frac{1-F_i(v_i)}{f_i(v_i)}=c$, which is $\overline{v}_j+v_i^*(\overline{v}_j)-\alpha_i\frac{1-F_i(v_i)}{f_i(v_i)}=c$ and leads to $(v_i+v_j^*(v_i)-c)-\frac{1-F_i(v_i)}{f_i(v_i)}\leq 0$. Note that $v_j^*(v_i)=c-v_i+\alpha_i\frac{1-F_i(v_i)}{f_i(v_i)}+\alpha_j\frac{1-F_j\left(v_j^*(v_i)\right)}{f_j\left(v_j^*(v_i)\right)}$. Thus, the expression $(v_i+v_j^*(v_i)-c)-\frac{1-F_i(v_i)}{f_i(v_i)}$ can be written as $\alpha_j\frac{1-F_j\left(v_j^*(v_i)\right)}{f_j\left(v_j^*(v_i)\right)}-(1-\alpha_i)\frac{1-F_i(v_i)}{f_i(v_i)}$. The derivative is

$$\begin{aligned}&\frac{\partial\left(\alpha_j\frac{1-F_j\left(v_j^*(v_i)\right)}{f_j\left(v_j^*(v_i)\right)}-(1-\alpha_i)\frac{1-F_i(v_i)}{f_i(v_i)}\right)}{\partial v_i}\\ =\;&\alpha_j\frac{\partial v_j^*}{\partial v_i}\frac{\partial\left(\frac{1-F_j\left(v_j^*(v_i)\right)}{f_j\left(v_j^*(v_i)\right)}\right)}{\partial v_j}-(1-\alpha_i)\frac{\partial\left(\frac{1-F_i(v_i)}{f_i(v_i)}\right)}{\partial v_i}\end{aligned}$$

which is always positive if $\frac{1-F_j(v_j)}{f_j(v_j)}$ is decreasing in v_j and $\frac{1-F_i(v_i)}{f_i(v_i)}$ is decreasing in v_i. In this case $(v_i+v_j^*(v_i)-c)-\frac{1-F_i(v_i)}{f_i(v_i)}$ is monotone in v_i and $\frac{\partial g(v_j^*(v_i))}{\partial v_i}$ switches sign only once. $\square$

It is left to show that $[-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))]$ is weakly positive. We can write

$$\begin{aligned}
&[-g(v_j) + (v_j + v_i^*(v_j) - c)(1 - F_i(v_i^*(v_j))(1 - F_j(v_j))] \\
=&\int_{v=v_j^*(\overline{v}_i)}^{v_j} (1 - F_j(v_j)) \left[(v_i^*(v_j) + v_j - c) f_i\left(v_i^*(v_j)\right) - \left(1 - F_i\left(v_i^*(v_j)\right)\right)\right] \frac{\partial v_i^*(v_j)}{\partial v_j} dv_j \\
&+ \int_{v=v_j^*(\overline{v}_i)}^{v_j} [(1 - F_i\left(v_i^*(v_j)\right))(v_i^*(v_j) + v_j - c)(-f_j(v_j) \\
&+ (v_i^*(v_j) + v_j - c)(1 - F_j(v_j))(-f_i\left(v_i^*(v_j)\right))\frac{\partial v_i^*(v_j)}{\partial v_j} \\
&+ (1 - F_i\left(v_i^*(v_j)\right))\left(1 + \frac{\partial v_i^*(v_j)}{\partial v_j}\right)(1 - F_j(v_j))]dv_j \\
=&\int_{v=v_j^*(\overline{v}_i)}^{v_j} (1 - F_i\left(v_i^*(v_j)\right))(v_i^*(v_j) + v_j - c)(-f_j(v_j)) + (1 - F_i\left(v_i^*(v_j)\right))(1 - F_j(v_j))dv_j \\
=&-\int_{v=v_j^*(\overline{v}_i)}^{v_j} \left(1 - F_i\left(v_i^*(v_j)\right)\right) \left[(v_i^*\left(v_j\right) + v_j - c) f_j(v_j) - (1 - F_j(v_j))\right] dv_j \qquad (2.16)
\end{aligned}$$

The argument to show (15) is positive if the hazard rates are monotone is analogous to the argument before for $g(v_j^*(v_i)) \geq 0$ because we can write

$$g(v_j^*(v_i)) = -\int_{v=v_i^*(\overline{v}_j)}^{v_i} (1 - F_j(v_j^*(v_i))) \left[(v_i + v_j^*(v_i) - c) f_i(v_i) - (1 - F_i(v_i))\right] dv_i$$

which is symmetric to (15). □

Chapter 3

Bayesian Persuasion with Private Experimentation[1]

3.1 Introduction

This paper studies a situation in which an agent tries to persuade a principal by providing experimental evidence. Having a similar motivation Kamenica and Gentzkow (2011) show how an agent optimally tries to persuade a Bayesian principal if he can design a hard (i.e., non-manipulable) signal where the realization is publicly observable. Often, however, information acquisition occurs in private. In this paper the agent may privately run as many experiments as desired and selectively reveal as many results as he wants. We assume that an experimental outcome is hard evidence, that the agent designs the precision of each experiment contingent on the experimentation history and that his decision to continue searching also depends on the experimentation history.

Our first contribution is to compare the persuasion probability in the sender preferred equilibrium under private and public experimentation. If under private experimentation the agent could use the same signal structure for persuasion as in Kamenica and Gentzkow (2011), then he would excessively search for favorable evidence if experimentation costs are not too high. Continuing the search can be profitable even if the agent observes unfavorable outcomes, as there is a chance that future experiments yield persuasive evidence. Naturally, he would not reveal any unfavorable outcomes. Excessive private experimentation and selective information revelation dilutes the value of revealed evidence such that one outcome that is generated with an experiment that has the same structure as in Kamenica and Gentzkow (2011) may not

[1]This chapter is joint work with Mike Felgenhauer.

be persuasive. However, there are equilibria that yield the same persuasion probability as in Kamenica and Gentzkow (2011), but in these equilibria the agent has to run multiple experiments for persuasion. This suggests a trade-off: If the agent cares sufficiently about a favorable decision, then he might be willing to obtain a higher persuasion probability at the expense of running several costly experiments.

Yet, this is not the case. In the sender preferred equilibrium the agent runs just one experiment on the equilibrium path. To counteract the incentive for excessive private experimentation the agent designs this experiment with a sufficiently high precision. If the outcome now is unfavorable, then it is too unlikely to find favorable evidence to justify the costs of further experimentation. A sufficiently high precision of the acquired signal deters excessive private experimentation, but it lowers the persuasion probability. This equilibrium has a surprisingly simple structure. The precision of the single experiment is such that it yields a positive outcome with certainty in the state in which the principal prefers the same decision as the agent. In the state in which their preferences differ, the precision is such that it just does not pay for the agent to continue experimentation if he knows this state. In the sender preferred equilibrium the agent does not run multiple experiments in order to obtain a higher persuasion probability, even if he cares substantially about a favorable decision. The persuasion probability in the sender preferred equilibrium under private experimentation is lower than under public experimentation.

We analyze the players' payoffs in the sender preferred equilibrium further. The agent does not benefit from the option to experiment privately. He runs a single experiment as under public experimentation, but the persuasion probability is lower. The principal, on the other hand, enjoys an advantage from the agent's commitment problem. As the high precision of the revealed evidence in the sender preferred equilibrium is not diluted by further private experimentation, the principal obtains higher quality information than under public experimentation.

We finally show that the higher the agent's stakes, the better is the decision quality from the principal's perspective. The reason is that the more the agent cares about a favorable decision, the higher the precision of the unique experiment has to be in order to deter further experimentation in case of an adverse outcome. A higher signal precision improves the decision quality.

This paper is structured as follows. The next section reviews the related literature. Section 3.3 contains a description of the model. In section 3.4 we briefly revise Kamenica and Gentzkow's (2011) equilibrium analysis under public experimentation. Section 3.5 contains our equilibrium analysis of private experimentation. In this section we also compare our results to Kamenica and Gentzkow (2011) and provide a comparative statics analysis. Section 3.6 concludes. Omitted proofs can be found in the appendix.

3.2 Literature

Our paper is related to the literature on persuasion (e.g., Jovanovic, 1982, Milgrom and Roberts, 1986, Glazer and Rubinstein, 2001, 2004, 2006), the literature on strategic experimentation (e.g., Rothschild, 1974, Aghion et al., 1991, Bolton and Harris, 1999, Keller et al., 2005 and Rosenberg et al., 2007)[2] and the literature that combines both strands (e.g., Henry, 2009, Brocas and Carrillo, 2007, Gentzkow and Kamenica, 2012, 2014, Felgenhauer and Schulte, 2013, Kamenica and Gentzkow, 2011).

Henry (2009) and Brocas and Carrillo (2007) investigate private experimentation in settings where the principal knows or can deduce the number of experiments that the agent runs. Their models allow for an unraveling argument a la Milgrom and Roberts (1986). However, given that experimentation occurs in private and is sequential, we think that it is more natural that the decision to continue experimenting is history dependent and unobservable.[3] Sceptical beliefs a la Milgrom and Roberts are, in general, not helpful in such a setting: The principal, in general, cannot deduce the number of experiments that the agent ran, she only knows the equilibrium experimentation plan. Nevertheless, in our paper it turns out that the agent in the sender preferred equilibrium runs a single experiment and, hence, this experiment is basically public.

Gentzkow and Kamenica (2014) use the same setup as in Kamenica and Gentzkow (2011), but in contrast to their previous paper they assume that experimentation is costly. They show that the concavification approach of Kamenica and Gentzkow (2011) extends to settings where the costs of a signal are proportional to the expected reduction in uncertainty. Gentzkow and Kamenica (2012) study competition between several agents who try to persuade a principal. They find that the equilibrium outcomes are identical regardless of whether the principal

[2] For a survey see Bergemann and Valimaki (2008).

[3] For example, if the agent finds too many unfavorable results in the first experiments, then he knows that he cannot persuade the principal by conducting the remaining experiments. As experimentation is private, he cannot be forced to continue costly experimentation until the ex ante determined number of experiments is conducted.

observes the signal or whether the agent sends a message to the principal and, therefore, can decide to be vague about his signal.

Felgenhauer and Schulte (2013) also consider a situation in which an agent tries to persuade a principal by providing experimental evidence stemming from sequential private experimentation. They characterize the set of equilibria and show that as the agent's stakes increase sufficiently, then he has to provide more positive outcomes for persuasion in any equilibrium with persuasion. The key difference to the present paper is that there the precision of the experiments is exogenous and that the agent there prefers equilibria in which he just persuades the principal.[4] In contrast to Felgenhauer and Schulte (2013), our major contribution is to compare the properties (like the persuasion probability, the players' payoffs and the decision quality) of the sender preferred equilibrium under private experimentation (with endogenous precision) and under public experimentation a la Kamenica and Gentzkow (2011).

3.3 Model

3.3.1 Preferences

A principal chooses $x \in \{0, 1\}$. Her payoff depends on an unknown state of the world $s \in \{0, 1\}$, with $prob\{s = 1\} = \frac{1}{2}$. The principal's utility is

	$s = 1$	$s = 0$
$x = 1$	1	$1 - p_d$
$x = 0$	p_d	1

with $p_d \in (\frac{1}{2}, 1)$. The principal, thus, would like to match the decision x with the state of the world s if she knew s. At the optimum she only chooses $x = 1$ if her posterior belief passes the "threshold of doubt" p_d, i.e., the posterior that $s = 1$ must be weakly greater than p_d.

There is an agent who prefers $x = 1$ regardless of s. His prior belief is also $prob\{s = 1\} = \frac{1}{2}$. His gross utility is U if $x = 1$ and 0 otherwise. Experimentation costs have to be subtracted

[4] In the sender preferred equilibrium in Felgenhauer and Schulte (2013) the number of favorable outcomes required for persuasion is such that the principal's threshold of doubt is just passed. Here instead, the agent provides evidence with a high precision which stems from a single experiment and, thus, the principal's posterior is above the threshold of doubt. As a consequence, in the sender preferred equilibrium in Felgenhauer and Schulte (2013) the principal is *just* persuaded upon the provision of persuasive evidence regardless of the agent's stakes, whereas here, the endogenous precision implies that the quality of the information she receives strictly increases in the agent's stakes.

from gross utility.

3.3.2 Experimentation

The agent has access to an experimentation technology that can generate signals about s. He can run as many experiments as desired. $y_t \in \{0,1\}$ is the outcome of the t^{th} experiment. We call $y_t = 1$ a positive outcome and $y_t = 0$ an adverse outcome (of experiment t). The agent chooses the precision of an experiment t, i.e., $prob\{y_t = 1 \mid s = 1\} = p_t^1$ and $prob\{y_t = 0 \mid s = 0\} = p_t^0$, $p_t^i \in [0,1]$, $i \in \{0,1\}$. The signals from experiments are, therefore, conditionally independent. Define $p_t \equiv (p_t^1, p_t^0)$. Let $p_t^1 \geq 1 - p_t^0$, i.e., a positive outcome $y_t = 1$ is more likely if $s = 1$ than if $s = 0$. Denote with $h_t = \{(y_i, p_i)\}_{i=1,\ldots,t}$ the experimentation history after the first t experiments. For simplicity, we assume that running an experiment costs $c \geq 0$ regardless of the precision.[5] The choice of p_t is history dependent. The decision to continue running experiments or to stop searching is also history dependent.

The agent cannot manipulate or make up experimental outcomes, i.e., y_t is "hard" information. He can reveal any subset of the acquired outcomes to the principal. Denote the reported results by $m = \{(y_i, p_i)\}_{i=1,\ldots,K}$, where y_i is the outcome of some experiment i that is revealed to the principal (where i does not refer to the time index t).[6] K is the finite number of outcomes contained in the message. The principal observes message m but she cannot observe the experimentation history. The agent cannot prove that he did not conduct a particular experiment.

3.3.3 Timing

First, there is an experimentation phase, which we model as a time interval. At each point of time within this phase the agent may conduct an experiment. This implies that if he runs an experiment, then he may still conduct as many experiments as desired before the phase ends. Therewith, we exclude the possibility of inferring information from the length of the experimentation phase.[7] After the experimentation phase the agent sends message m and

[5] With increasing costs standard arguments apply.

[6] The principal observes the precision of an experiment, once the outcome of this experiment is presented. This assumption is natural in many applications. E.g., if a theoretical argument is considered as evidence, then a scientific audience, e.g., referees, editors or seminar participants, can assess its quality.

[7] Hopenhayn and Squintani (2011) study in a different context the case where the decision maker can deduce something from the time elapsed until he receives information. A longer period may, e.g., suggest many failed experiments or that the agent ran a complex experiment. Our model abstracts from these issues. Often experiments differ regarding the time they require until completed and it is difficult to deduce information from the time elapsed.

then the principal chooses x.

3.3.4 Strategies and equilibrium concept

The agent's strategy consists of an experimentation plan and an announcement plan. For each experimentation history the experimentation plan specifies whether to continue or to stop experimenting and which precision to use. The announcement plan states what to reveal to the principal after stopping. A strategy for the principal is to choose $x \in \{0, 1\}$ for each possible announcement.

The equilibrium concept we use is weak perfect Bayesian equilibrium. Due to the power of off the equilibrium path beliefs there are many equilibria. For example, the principal may have beliefs that force the agent to provide high quality evidence by believing that the agent privately ran sufficiently many high quality experiments with unfavorable outcomes if other (relatively) low quality evidence is presented. In order to limit the power of off the equilibrium path beliefs, we focus on sender preferred equilibrium. An equilibrium is sender preferred if there is no other equilibrium in which the agent is strictly better off. This refinement makes our results easily comparable to Kamenica and Gentzkow (2011) who also study sender preferred equilibria.

3.4 Persuasion a la Kamenica and Gentzkow (2011)

Kamenica and Gentzkow assume that the agent runs a single public experiment. In order to maximize the persuasion probability he chooses $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$.[8] The intuition for $p_1^1 = 1$ is that the agent does not want to get an adverse outcome if $s = 1$. He also wants to get a positive outcome when the state is bad, but he has to choose p_1^0 such that he can still persuade the principal. Therefore, he maximizes $prob\{y_1 = 1 \mid s = 0\}$ subject to $prob\{s = 1 \mid y_1 = 1\} \geq p_d$. The persuasion probability is maximal if $prob\{s = 1 \mid y_1 = 1\} = p_d$, i.e., such that the principal is indifferent between $x = 1$ and $x = 0$ upon the observation of a positive outcome, yielding $p_1^0 = \frac{2p_d-1}{p_d}$. The persuasion probability is $\frac{1}{2p_d}$. The precision $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$ does not depend on the agent's stakes U/c.

[8]Kamenica and Gentzkow assume $c = 0$. If experimentation is costly, then starting experimentation is profitable if $U/c \geq 2p_d$. Otherwise, experimentation is too costly.

3.5 Private experimentation

Each equilibrium is characterized by the set of messages M^* which can persuade the principal. This set results from what the principal believes about the experiments which the agent does not reveal to her. An equilibrium with information provision has to satisfy the following two conditions. First, the agent has to have an incentive to start experimentation. Second, the principal is persuaded if his posterior that $s = 1$ given the provided evidence passes the threshold of doubt, i.e., $prob\{s = 1 \mid m^*\} \geq p_d$ has to be satisfied, where $m^* \in M^*$ denotes a message that persuades the principal in equilibrium. Otherwise, she chooses $x = 0$. We assume that the principal's beliefs are such that the agent can always persuade the principal if he provides at least one positive outcome with precision $p_i^0 = 1$. A message m^* containing such an outcome implies $prob\{s = 1 \mid m^*\} = 1$.

In contrast to public experimentation, $prob\{s = 1 \mid m^*\}$ and, hence, whether the announcement can persuade the principal depends on U/c. In general, persuasion in equilibrium requires that the agent sometimes stops experimenting unsuccessfully. Otherwise, any message that does not reveal a state with certainty is uninformative with respect to s.

Remark 3.1. Consider $U/c \in [0, \frac{p_d}{1-p_d}]$. The agent's behavior in the sender preferred equilibrium under private experimentation on the equilibrium path is the same as under public experimentation.

If U/c is too small, i.e., $U/c \in [0, 2p_d)$, then there is no equilibrium with persuasion under both schemes. The probability to obtain U is not high enough for the agent to justify the costs of running an experiment. If $U/c \in [2p_d, \frac{p_d}{1-p_d}]$, then there is an equilibrium in which the agent on the equilibrium path runs a single experiment and this experiment has the same precision as under public experimentation and the principal is persuaded if the agent presents a corresponding positive outcome. It follows that for $U/c \in [0, \frac{p_d}{1-p_d}]$, the decision quality, the persuasion probability, etc. are the same under public and private experimentation. In the following we focus on $U/c > \frac{p_d}{1-p_d}$.

Remark 3.2. Consider $U/c > \frac{p_d}{1-p_d}$. There is no equilibrium where the agent on the equilibrium path runs one experiment with precision $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$, i.e., the same precision as of the experiment under public experimentation, and no further experiments.

In order to verify the remark, suppose instead that there is such an equilibrium. If $y_1 = 0$, the agent knows that $s = 0$. It is worthwhile to continue experimenting, given that the principal can be persuaded with one positive outcome from an experiment with precision $(1, \frac{2p_d-1}{p_d})$, if $U/c > \frac{1}{(1-p_1^0)} = \frac{p_d}{1-p_d}$. In this case upon the provision of a positive outcome, the probability that $s = 1$ must be below the threshold of doubt p_d, since the positive outcome can stem from many experiments. Consequently, this cannot be an equilibrium if U/c is sufficiently high.

If U/c is sufficiently high, i.e., $U/c > \frac{p_d}{1-p_d}$, then there is no equilibrium in which the agent runs a single experiment that leads to the maximum persuasion probability as under public experimentation. The agent in equilibrium, however, may run multiple experiments and may reveal more than one outcome.

We now describe equilibria that yield the same persuasion probability as under public experimentation, but where the agent runs more than one experiment. Consider $U/c \in [2p_d + 2N, \frac{p_d}{1-p_d} + 2N]$ with $N \geq 1$.[9] There exists a weak perfect Bayesian equilibrium with the following properties. The principal is persuaded if the agent presents a positive outcome from an experiment with precision $(1, \frac{2p_d-1}{p_d})$, i.e., the same precision as under public experimentation, and in addition N positive outcomes, where the latter stem from experiments that are uninformative regarding s. The number of positive outcomes from uninformative experiments N is sufficiently low such that the agent starts searching with an experiment with $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$ and sufficiently high such that he stops searching unsuccessfully if $y_1 = 0$. If $y_1 = 1$, then he runs uninformative experiments until he finds the required N further positive outcomes. As the principal knows that the agent runs only one informative experiment and this experiment has precision $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$, she is just persuaded upon the provision of the $N + 1$ required pieces of evidence.[10]

It is therefore possible to obtain the maximum persuasion probability, but the agent has to run multiple experiments for persuasion. This suggests a trade-off between the persuasion probability and the number of experiments from the agent's perspective. If the agent cares sufficiently about a favorable decision relative to the costs per experiment, then we might conjecture that in the sender preferred equilibrium the maximum persuasion probability can be obtained.

In the following we show that private experimentation limits the extent to which persuasion is possible in the sender preferred equilibrium. Proposition 3.1 asserts that the agent in

[9] According to Remark 3.2, there is no equilibrium in which the agent runs one experiment a la Kamenica and Gentzkow if $U/c > \frac{p_d}{1-p_d}$ and no further experiment. Later we often refer to such parameters $U/c > \frac{p_d}{1-p_d}$. Note that for each $p_d \in (\frac{1}{2}, 1)$, there is an N, such that $2p_d + 2N > \frac{p_d}{1-p_d}$.

[10] In the appendix, in part "verification of Claim 1", we show that the equilibrium conditions are met.

the sender preferred equilibrium runs a single experiment with $p_1^1 = 1$ (as in Kamenica and Gentzkow, 2011), but with p_1^0 that maximizes the persuasion probability subject to the constraint that further experimentation after an initial adverse outcome is deterred.

Proposition 3.1. *Consider $U/c > \frac{p_d}{1-p_d}$. (i) There is an equilibrium in which the agent on the equilibrium path runs one experiment with precision $(p_1^1, p_1^0) = (1, \frac{U-c}{U})$ and no further experiment. (ii) The equilibrium in (i) is sender preferred.*

We can now compare the persuasion probabilities and the decision quality under public and private experimentation in the sender preferred equilibria. Our interpretation of the decision quality is motivated by the principal's preferences. She wants to match the state with the decision. The principal is better off (and we say that the decision quality increases) if either $prob\{x = 0 \mid s = 1\} \downarrow$ without changing $prob\{x = 1 \mid s = 0\}$ or $prob\{x = 1 \mid s = 0\} \downarrow$ keeping $prob\{x = 0 \mid s = 1\}$ constant or both decrease.

Consider a finite $U/c > \frac{p_d}{1-p_d}$. Under both schemes $p_1^1 = 1$. $p_1^1 = 1$ implies that the principal does not make the "wrong" decision $x = 0$ in state $s = 1$. Therefore, $prob\{x = 0 \mid s = 1\}$ is the same under public and private experimentation and in this state both schemes are equally good for the principal. In order to deter a second experiment after $y_1 = 0$, p_1^0 has to be greater under private experimentation if U/c is sufficiently high. A greater p_1^0 implies that the principal makes the wrong decision $x = 1$ in state $s = 0$ with a lower probability. Therefore, $prob\{x = 1 \mid s = 0\}$ is lower under private experimentation than under public experimentation. As $prob\{x = 0 \mid s = 1\}$ is the same under both schemes and $prob\{x = 1 \mid s = 0\}$ is lower under private experimentation, the decision quality is higher under private experimentation. The same p_1^1 under both schemes and a greater p_1^0 under private experimentation also implies that the persuasion probability is lower if search is private.

Consider next $c = 0$ (as in Kamenica and Gentzkow, 2011) or $U \to \infty$. In these cases, if there is a hypothetical equilibrium in which the agent can find with a positive probability evidence that has the power to persuade the principal, even if the agent knows that $s = 0$, then he searches for it until he finds it. But then the evidence is uninformative with respect to the state of the world s. The principal should not be persuaded by such evidence, yielding a contradiction to the notion of an equilibrium. However, the principal can always be persuaded in equilibrium if the agent reveals a positive outcome of an experiment that perfectly mirrors the state of the world.[11] In the sender preferred equilibrium the persuasion probability is 1/2 and the decision matches the state.[12] It follows that the decision quality is greater under private experimentation, but the persuasion probability is lower than under the public scheme.

The following proposition summarizes these results.

Proposition 3.2. *Consider $U/c > \frac{p_d}{1-p_d}$. (i) The persuasion probability in the sender preferred equilibrium under private experimentation is lower than under public experimentation. (ii) The decision quality under private experimentation is higher.*

The persuasion probability under public experimentation is $\frac{1}{2p_d}$. The persuasion probability under private experimentation is $\frac{U+c}{2U}$ if $U/c > \frac{p_d}{1-p_d}$ is finite and it is 1/2 if $U/c \to \infty$. Under both schemes the agent runs a single experiment, but the persuasion probability is higher under public experimentation, rendering private experimentation less attractive for the agent.

The principal on the other hand benefits from the agent's commitment problem, as $prob\{x = 1 \mid s = 0\}$ is lower under private experimentation and $prob\{x = 0 \mid s = 1\} = 0$ under both schemes.

Corollary 3.1. *Consider $U/c > \frac{p_d}{1-p_d}$. The agent strictly prefers public to private experimentation in the sender preferred equilibrium. The principal strictly prefers private to public experimentation in the sender preferred equilibrium.*

Finally, we analyze what happens in the sender preferred equilibrium under private experimentation when the stakes of the agent U/c change. The precision of experiment i increases if ceteris paribus $p_i^1 \uparrow$ or $p_i^0 \uparrow$ or both increase. In the sender preferred equilibrium $p_1^1 = 1$ irrespective of U/c. The stakes of the agent U/c exclusively affect p_1^0. If U/c increases, then p_1^0 has to adjust such that the agent does not have an incentive to run a second experiment if he observes $y_1 = 0$, i.e., p_1^0 has to increase.

Proposition 3.3. *Consider $U/c > \frac{p_d}{1-p_d}$. (i) The precision of the experiment that is run on the equilibrium path increases in U/c. (ii) The persuasion probability decreases in U/c, (iii) The decision quality increases in U/c.*

The persuasion probability decreases in U/c, but the agent's experimentation costs decrease if $c \downarrow$ and in case he obtains his preferred decision, then his payoff increases if $U \uparrow$. The

[11] If $p_1^0 = 1$ and $y_1 = 0$, then the agent knows that $s = 0$. In this case running a further experiment i with $p_i^0 = 1$ must yield $y_i = 0$, which does not persuade the principal and is futile.

[12] In the appendix, part "verification of Claim 2", we show that in any equilibrium the maximum persuasion probability is 1/2 if $U/c \to \infty$.

agent's ex ante utility in the sender preferred equilibrium under private experimentation is $\frac{1}{2}(U-c)+\frac{1}{2}((1-\frac{U-c}{U})U-c)$, which is equivalent to $\frac{1}{2}(U-c)$.

If U/c increases, then the principal makes the wrong decision in state $s=0$ with a lower probability, since p_1^0 increases. In state $s=1$ on the other hand $prob\{x=0 \mid s=1\}=0$ for all $U/c > \frac{p_d}{1-p_d}$, as $p_1^1=1$ regardless of the agent's stakes.

Corollary 3.2. *Consider $U/c > \frac{p_d}{1-p_d}$. The agent's ex ante utility increases in U/c in the sender preferred equilibrium under private experimentation. The principal's ex ante utility increases in the agent's stakes U/c in the sender preferred equilibrium under private experimentation.*

If, e.g., a politician is interpreted as the principal and the agent as a lobby, then our analysis suggests, that the quality of informational lobbying and, hence, the decision quality increase in the stakes of the lobby. A similar point can be made if a researcher with career concerns is viewed as the agent and the principal is an editor. Young researchers aspiring tenure may care more about a publication in a good journal than researchers with tenure. In order to get published in the same journal, the former may have to write higher quality papers.

3.6 Discussion

There is an abundance of situations in which arguments are exchanged (e.g., in lobbying, public discussions of economic policies, the academic publishing process, etc.). Arguments have an inherent meaning and are not cheap talk. We think that many arguments, like logical arguments or a regression analysis on a public database, can be viewed as hard and imperfect decision relevant evidence.[13] Furthermore, arguments often have to be acquired. In their elegant paper Kamenica and Gentzkow (2011) describe, e.g., a trial as a public experiment yielding an argument, where a prosecutor can design the error probabilities in order to achieve his objectives. The error probabilities can be affected, e.g., by structuring the examination of witnesses in court.

[13] Once a regression method is described and the database is public, manipulation is not possible. A similar case is made by Felgenhauer and Schulte (2013) for logical arguments. They interpret logical arguments as decision relevant hard information that result from experimentation: If the assumptions underlying a logical argument are revealed, then they cannot be manipulated. The deductions are logical and logic cannot be manipulated. Logical arguments have persuasive power, therefore, they can be viewed as signals about a decision relevant state of the world. The signals are imperfect, as the underlying assumptions do not cover every real world aspect. A thought experiment (i.e., drawing a set of assumptions and making a deduction) yields a signal. For such arguments the assumption that they are acquired in private by running a series of thought experiments and selectively revealed for persuasion is natural. Such arguments have persuasive power and, hence, they can be viewed as decision relevant. Naturally, they are also imperfect.

While a public experiment is an interesting case for generating an argument, there are many situations in which arguments stem from sequential private experimentation. For example, if an agent wants to persuade with logical arguments, then he runs a series of thought experiments. He privately chooses the properties of each experiment, e.g., by choosing the conceptual framework from which to draw a set of specific assumptions. The outcome of each thought experiment is privately observed. Naturally, he may run as many thought experiments with properties of his choice (depending on what he has learnt from previous experimentation) as desired and selectively reveal the results. If an agent wants to persuade with an empirical analysis using a public database, then he may run a series of regressions. By choosing the econometric method and the model specification he chooses the properties of each experiment. Again search is sequential and the results are revealed selectively.

If arguments stem from sequential private experimentation and are selectively revealed, then the revealed evidence should not be taken at face value. The value of such arguments depends on the equilibrium experimentation plan, which in turn is influenced by experimentation costs and the agent's benefit from a favorable decision. In our paper an agent wants to persuade a principal and we focus on sender preferred equilibria. An experimentation plan is a complex object, since the agent in our model can make many history dependent choices. The agent also has considerable degrees of freedom regarding the messages that he can send, as he may reveal any subset of the acquired evidence, including "counterarguments". We show that a sender preferred equilibrium has a surprisingly simple structure. In equilibrium the agent runs a single experiment. This experiment correctly predicts the state in the state where the players prefer the same decision. In the state where the preferences differ the experiment is such that the agent is just deterred from continuing the search. The simplicity of this equilibrium may prove to be useful in other applications. In this equilibrium the persuasion probability is shown to be lower than under public experimentation, since the experiment has a higher precision.

The number of experiments that are run under public and private experimentation in the sender preferred equilibrium are the same, but the persuasion probability under the latter is lower. Consequently, the agent is worse off under private experimentation. The principal benefits from the agent's commitment problem not to run further private experiments and, therefore, she prefers private experimentation. As an application consider a pharmaceutical company (the agent) that attempts to persuade the U.S. Food and Drug Administration (the principal) to approve a newly developed drug. Given the enormous R&D costs in the pharmaceutical industry, it is plausible that the company prefers that a new drug is approved, even if its merits are doubtful. The FDA instead would like to make the "appropriate" decision, which could be against the company. The FDA mainly has to rely on tests, e.g., clinical studies, provided by the company, which in turn has an incentive to behave strategically. The decision quality can be influenced by the rules under which evidence can be acquired and revealed and what evidence is permitted to be considered as decision relevant. The evidence production may be designed as public, by imposing severe penalties if this rule is violated, or as private. Our paper suggests that, even in the sender preferred equilibrium, the FDA would be better off under the private scheme, but the company would benefit more from public experimentation.

We finally show that the persuasion probability decreases and the decision quality increases in the stakes of the agent. In a context where an interested party (like a lobby, student, researcher, etc.) tries to persuade a decision maker (like a politician, teacher, editor, etc.) to choose a favorable action (like a policy, a better mark in the exam, the publication of a paper, etc.) this means that the decision maker is better off the more the interested party cares about a favorable decision. The interested party then has to provide higher quality information in order to be able to commit not to run additional private experiments after an initial failure. If the decision maker can influence the experimentation costs, then she benefits from lowering costs by an analogous reasoning.

It is well established that inefficiencies occur in the presence of private information even in the simplest economic situations (e.g., Myerson and Satterthwaite, 1983). These inefficiencies are often derived in mechanisms in which only cheap talk messages are feasible. In future research it would be interesting to analyze the repercussions of an exchange of arguments on efficiency in a mechanism design framework with transfers and multiple agents.

3.7 Appendix

A Modified public experimentation

If the principal implements a public scheme in the sense that she can only be persuaded with evidence obtained from public experiments, the agent nevertheless may have the option to run experiments privately. Let us investigate a modified public scheme. Under this scheme, the agent can run private as well as public experiments, but he can exclusively persuade the principal with evidence obtained from public experiments. Analogous to our previous analysis of private experimentation, there are equilibria under the modified public scheme in which the agent has to conduct high quality public experiments in order to be able to persuade the principal. Such an equilibrium can be supported with similar beliefs as in the private scheme: If the principal off the equilibrium path observes low quality public experiments, then she should think that the agent also privately collected sufficiently adverse evidence such that it does not pay to run additional public experiments with the required quality. The corresponding beliefs induce a decision against the agent and, therefore, destroy the agent's incentive to run low quality public experiments.

Consider sender preferred equilibria as in the main part of the paper. The best equilibrium for the agent under the modified public scheme is where he runs exactly one public experiment with the same quality as under pure public experimentation, which leads to $prob\{s = 1|x = 1\} = p_d$. This equilibrium has the highest probability to persuade, and the agent incurs the cost of running only one experiment. Thus, the agent does not run any private experiment and the equilibrium is equivalent to the equilibrium under pure public experimentation. It follows that in the sender preferred equilibrium the principal prefers the pure private scheme to modified public experimentation and the agent prefers modified public experimentation to the pure private scheme.

B Proofs

Proof of Remark 3.1: Consider $U/c \in [2p_d, \frac{p_d}{1-p_d}]$. There is an equilibrium in which the principal is persuaded if the agent provides a single positive outcome and this outcome stems from an experiment with precision $(1, \frac{2p_d-1}{p_d})$. In this equilibrium the agent on the equilibrium path runs one experiment and this experiment has precision $(1, \frac{2p_d-1}{p_d})$.[14] An equilibrium

[14]The full characterization of the equilibrium is analogous to the equilibrium described in the proof of Proposition 3.1 (i).

condition is that the agent has to start experimenting. His ex ante utility from running a single experiment (anticipating persuasion if $y_1 = 1$ and unsuccessful stopping if $y_1 = 0$) with $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$ is $\frac{1}{2}(U-c)+\frac{1}{2}((1-\frac{2p_d-1}{p_d})U-c)$. This utility is non-negative if $U/c \geq 2p_d$. A further equilibrium condition is that the agent has an incentive to stop experimenting if $y_1 = 0$. It is not worthwhile to continue experimenting, given that the principal can be persuaded with one positive result of the same precision, if $U/c \leq \frac{1}{(1-p_1^0)} = \frac{p_d}{1-p_d}$. The principal in equilibrium knows that the agent runs an experiment with precision $(p_1^1, p_1^0) = (1, \frac{2p_d-1}{p_d})$ and no further experiment. Hence, the decision rule is optimal as verified in the section on public experimentation. This equilibrium is sender preferred, as the agent runs a single experiment with the maximum persuasion probability. Consider $U/c \in [0, 2p_d)$. In this case there is no equilibrium with persuasion, as it does not pay to start experimenting even if the agent could persuade with the minimum number of experiments required for persuasion and the maximum persuasion probability. Q.E.D.

Verification of Claim 1: Consider a weak perfect Bayesian equilibrium in which the principal is persuaded if the agent provides N positive outcomes from uninformative experiments and one positive outcome from an experiment with precision $(1, \frac{2p_d-1}{p_d})$ if $U/c \in [2p_d + 2N, \frac{p_d}{1-p_d} + 2N]$ as described in the text.[15] The number of experiments to be run until the N uninformative positive results are found follows a negative binomial distribution with success probability $\frac{1}{2}$ in state $s = 1$ and success probability $\frac{1}{2}$ in state $s = 0$. With success probability π, the expected number of experiments to be conducted until N positive outcomes are obtained is $\frac{N}{\pi}$. Ad interim the probability that $s = 1$ (and, therefore, the success probability is $\frac{1}{2}$) is $prob\{s = 1 \mid h_t\} \equiv \overline{p}$. With probability $1-\overline{p}$ the state is $s = 0$ (and, therefore, the success probability is $\frac{1}{2}$). Hence, the interim expected number of experiments to be run until the N positive outcomes are found is $\overline{p}\frac{N}{\frac{1}{2}} + (1-\overline{p})\frac{N}{\frac{1}{2}} = 2N$ regardless of $\overline{p}$.

An equilibrium condition is that the agent stops unsuccessfully if $y_1 = 0$. If $y_1 = 0$, the agent knows that $s = 0$, due to $p_1^1 = 1$. The probability that he finds a positive outcome from running a further experiment with precision $(1, \frac{2p_d-1}{p_d})$ is $1 - \frac{2p_d-1}{p_d}$. If he finds a positive

[15] The full characterization of the equilibrium is analogous to the equilibrium described in the proof of Proposition 3.1 (i) with one modification. There is a threshold posterior that can be greater than zero. The agent continues searching after off the equilibrium path histories h_t that do not contain a positive outcome from an experiment with precision $(1, \frac{2p_d-1}{p_d})$ iff the posterior implied by h_t is above the threshold posterior. In this case the next experiment has precision $(1, \frac{2p_d-1}{p_d})$. If the outcome of this experiment is positive, then the agent searches for the positive outcomes of the uninformative experiments until he finds the required number for persuasion. If the outcome of this experiment is adverse, then the agent stops searching unsuccessfully.

outcome from such an experiment, then he would continue searching for the positive outcomes from the uninformative experiments until he finds the required number of them, implying expected costs $2Nc$ (for the uninformative experiments). Hence, his expected utility from searching further in case $y_1 = 0$ is $(U - 2Nc)(1 - \frac{2p_d-1}{p_d}) - c$ and the agent stops if this expression is weakly below 0. This yields $U/c \leq \frac{p_d}{1-p_d} + 2N$.

A further equilibrium condition is that the agent starts searching. The ex ante expected utility is $\frac{1}{2}(U - 2Nc) + \frac{1}{2}((U - 2Nc)(1 - \frac{2p_d-1}{p_d})) - c$, which is equivalent to $\frac{U-2Nc}{2p_d} - c$. He has an incentive to start experimenting if $\frac{U-2Nc}{2p_d} - c \geq 0$, i.e., if $U/c \geq 2p_d + 2N$.

Finally, the agent has to have an incentive to continue searching until he finds the required $N+1$ positive outcomes for persuasion given that $y_1 = 1$ and he has found τ positive outcomes from uninformative experiments, with $\tau \in [0, N)$. In each such a contingency his stock of evidence that he can use for persuasion is greater than ex ante. Therefore, if the agent starts searching, then he also continues searching in each such contingency. Q.E.D.

Proof of Proposition 3.1: (i) Consider the following equilibrium.

Principal's strategy: The principal chooses $x = 1$ iff the agent sends a message $m^* \in M^*$, where one $m^* \in M^*$ contains a single outcome $y_i = 1$ with $(p_i^1, p_i^0) = (1, \frac{U-c}{U})$ of an experiment i (and no further outcome) and all other $m^* \in M^*$ contain at least one outcome $y_i = 1$ with $(p_i^1, p_i^0) = (p^1, 1)$, with some $p_1 \in (0, 1]$.

Experimentation plan: Given experimentation history h_t the agent:

- runs a further experiment with $(p_{t+1}^1, p_{t+1}^0) = (1, \frac{U-c}{U})$ if $prob\{s = 1 \mid h_t\} > 0$ and if h_t does not contain evidence such that sending a message $m^* \in M^*$ is feasible,
- stops searching if $prob\{s = 1 \mid h_t\} = 0$,
- stops searching if h_t contains evidence such that sending a message $m^* \in M^*$ is feasible.

Beliefs: On the equilibrium path beliefs are formed in accordance with Bayes' Law. Off the equilibrium path beliefs are such that upon the provision of a message $m' \notin M^*$ the principal forms a probability assessment over experimentation histories such that the probability that $s = 1$ conditional on this assessment is below the threshold of doubt. E.g., he may believe that the agent privately ran a single additional experiment i with $p_i^1 = p_i^0 = 1$ and $y_i = 0$.

We now show that the above constitutes an equilibrium:

Agent behavior: The agent anticipates that he can only persuade the principal with a $m^* \in M^*$. As the persuasion probability with precision $(p^1, 1)$, with $p^1 \in (0, 1]$, is (weakly) lower than with precision $(1, \frac{U-c}{U})$ for any h_t, the agent does not run an experiment with $(p^1, 1)$.

If h_t is such that sending a message $m^* \in M^*$ is feasible, then the agent stops searching successfully.

If h_t does not contain evidence such that a message $m^* \in M^*$ is feasible and $prob\{s = 1 \mid h_t\} = 0$, then he does not run a further experiment with $(p^1_{t+1}, p^0_{t+1}) = (1, \frac{U-c}{U})$ if $(1 - \frac{U-c}{U})U - c \leq 0$, which is satisfied as the agent is indifferent to continue.

If h_t does not contain evidence such that a message $m^* \in M^*$ is feasible and $prob\{s = 1 \mid h_t\} > 0$, then the agent has a strict incentive to continue searching.

As the agent is indifferent to continue searching if he knows $s = 0$, he has a strict incentive to start searching given his prior belief.

The agent on the equilibrium path runs a single experiment with $(p^1_1, p^0_1) = (1, \frac{U-c}{U})$ as $y_1 = 1$ is persuasive and $y_1 = 0$ yields $prob\{s = 1 \mid y_1\} = 0$.

Principal behavior: The principal knows that the agent runs a single experiment and this experiment has precision $(p^1_1, p^0_1) = (1, \frac{U-c}{U})$. She is persuaded by the message $m^* \in M^*$ that contains a single experimental outcome $y_i = 1$ with $(p^1_i, p^0_i) = (1, \frac{U-c}{U})$ if $\frac{p^1_i}{p^1_i + (1-p^0_i)} \geq p_d$, i.e., if $\frac{1}{1+(1-\frac{U-c}{U})} \geq p_d$, which is satisfied if $U/c > \frac{p_d}{1-p_d}$.

Any other $m^* \in M^*$ that contains at least one experimental outcome $y_i = 1$ with $(p^1_i, p^0_i) = (p^1, 1)$, with some $p_1 \in (0, 1]$, can also persuade her as such an outcome implies that $s = 1$ with probability 1 (regardless of the other elements of such a message).

Beliefs given an off the equilibrium path message $m' \notin M^*$ are such that $prob\{s = 1 \mid m'\} < p_d$ and, hence, in this case $x = 0$ is optimal.

(ii) Define $V_0(\overline{p})$ as the agent's expected utility if (I) $s = 0$, (II) he has not yet found an outcome that is an element of some $m^* \in M^*$, and (III) he continues experimenting according to his equilibrium experimentation plan given that he holds the posterior $prob\{s = 1 \mid h_t\} = \overline{p}$ and given that he has not yet found an outcome that is an element of some $m^* \in M^*$. Analogously, define $V_1(\overline{p})$ as his expected utility given $s = 1$.

The agent's utility if he has not yet found an outcome that he can use for persuasion at posterior $\overline{p}$ is

$$\overline{p}V_1(\overline{p}) + (1 - \overline{p})V_0(\overline{p}).$$

We use the following two lemmas in the proof.

Lemma 3.1. *An equilibrium with persuasion requires that the agent stops searching unsuccessfully at some posterior $\overline{p} < \frac{1}{2}$ if he has not yet found an outcome that is part of some*

$m^* \in M^*$.

Proof: An equilibrium with persuasion requires that the agent stops searching unsuccessfully at some posterior $\overline{p} < \frac{1}{2}$. If he instead in a hypothetical equilibrium only stops unsuccessfully if $\overline{p} \geq \frac{1}{2}$, then unsuccessful search implies that the posterior from the principal's perspective is above 1/2. Bayesian plausibility then requires that the average posterior from the principal's perspective after successful search is below 1/2. The principal's posterior may depend on the revealed evidence but at least for some $m^* \in M^*$ we have $prob\{s = 1 \mid m^*\} < 1/2 < p_d$. But then the principal should not be persuaded by such a message yielding a contradiction.

The agent is better off at a posterior $\overline{p}$ if he has a stock of evidence that he can use for persuasion than if he does not have such evidence. Hence, if the agent continues searching at posterior $\overline{p}$ if he does not have evidence that is part of some message $m^* \in M^*$, then he also continues searching at posterior $\overline{p}$ if he has a stock of evidence that he can use for persuasion. Q.E.D.

Lemma 3.2. *In an equilibrium with persuasion we have* $V_1(\frac{1}{2}) \geq 0$ *and* $V_0(\frac{1}{2}) \leq 0$ *and* $\frac{1}{2}V_1(\frac{1}{2}) + \frac{1}{2}V_0(\frac{1}{2}) \geq 0$.

Proof: $\frac{1}{2}V_1(\frac{1}{2}) + \frac{1}{2}V_0(\frac{1}{2}) \geq 0$ has to be satisfied, as the agent has to have an incentive to start experimenting.

We cannot have $V_1(\frac{1}{2}) < 0$ and $V_0(\frac{1}{2}) \leq 0$, as this violates $\frac{1}{2}V_1(\frac{1}{2}) + \frac{1}{2}V_0(\frac{1}{2}) \geq 0$. Similarly, we cannot have $V_1(\frac{1}{2}) \leq 0$ and $V_0(\frac{1}{2}) < 0$.

We cannot have $V_1(\frac{1}{2}) \geq 0$ and $V_0(\frac{1}{2}) > 0$. In this case the agent could (possibly suboptimally) at each posterior $\overline{p}$ replicate the same behavior as at the prior where he has not yet run an experiment, yielding $V_1(\frac{1}{2})$ and $V_0(\frac{1}{2})$ in the respective states with this modified plan at posterior $\overline{p}$.[16] As $\overline{p}$ only allocates probability mass to $V_1(\frac{1}{2})$ and to $V_0(\frac{1}{2})$ in $\overline{p}V_1(\frac{1}{2}) + (1-\overline{p})V_0(\frac{1}{2})$ (and both are greater than zero), the agent's benefit from continuing to search with the modified plan is greater than zero. With the optimal plan he is weakly better off. Therefore, he would not stop searching unsuccessfully for any posterior given that he has

[16] "Replicating" behavior means the following. Consider a history $h_t = \{(y_i, p_i)\}_{i=1,...,t}$, where the agent has not yet found evidence that is part of some $m^* \in M^*$ and where he faces posterior $\overline{p} < \frac{1}{2}$. Consider further a history $h_{t'} = \{(y_i, p_i)\}_{i=1,...,t,...,t'}$, with $t' \geq t$. Based on these histories construct an artificial history h_z as follows: Each (y_i, p_i) in h_z is equal to (y_{t+i}, p_{t+i}) in $h_{t'}$, with $i = 1, ..., z$ and $z = t' - t$.

The agent's experimentation plan specifies for each history h_t (i) whether the agent continues or stops experimenting and (ii) the precision of the next experiment if he continues experimenting. We say that the agent from history h_t on replicates the same behavior as at the prior where he has not yet found evidence that can be used for persuasion, if the agent facing history h_t runs the next experiment with precision p_1 and at each history $h_{t'} = \{(y_i, p_i)\}_{i=1,...,t,...,t'}$, with $t' > t$, he continues experimenting according to his experimentation plan, as if he faces history h_z instead of history $h_{t'}$.

not yet acquired an outcome that he can use for persuasion. Due to Lemma 3.1 this implies that he never stops searching unsuccessfully for all posteriors $\overline{p} < \frac{1}{2}$, which contradicts the notion of an equilibrium with persuasion.

We cannot have $V_1(\frac{1}{2}) \leq 0$ and $V_0(\frac{1}{2}) > 0$. In this case the agent could (possibly suboptimally) at each posterior $\overline{p} < \frac{1}{2}$ replicate the same behavior as at the prior where he has not yet run an experiment. In this case for any $\overline{p} < \frac{1}{2}$, more probability mass would be allocated to $V_0(\frac{1}{2})$ than ex ante and less probability mass would be allocated to $V_1(\frac{1}{2})$ than ex ante. I.e., if $\frac{1}{2}V_1(\frac{1}{2}) + \frac{1}{2}V_0(\frac{1}{2}) \geq 0$, then we have $\overline{p}V_1(\frac{1}{2}) + (1-\overline{p})V_0(\frac{1}{2}) > 0$ for all $\overline{p} < \frac{1}{2}$. Optimal behavior implies a weakly greater utility. Due to Lemma 3.1 this implies that he never stops searching unsuccessfully for all posteriors $\overline{p} < \frac{1}{2}$, which contradicts the notion of an equilibrium with persuasion.

Therefore, $\frac{1}{2}V_1(\frac{1}{2}) + \frac{1}{2}V_0(\frac{1}{2}) \geq 0$ requires $V_1(\frac{1}{2}) \geq 0$ and $V_0(\frac{1}{2}) \leq 0$. Q.E.D.

Call the equilibrium in which the agent on the equilibrium path runs exactly one experiment and this experiment has precision $(p_1^1, p_1^0) = (1, \frac{U-c}{U})$ "equilibrium SP" (where SP stands for "sender preferred"). The agent's ex ante utility in such an equilibrium is $\frac{1}{2}(U-c) + \frac{1}{2}((1 - \frac{U-c}{U})U - c)$, which is equivalent to

$$\frac{1}{2}(U-c). \tag{3.1}$$

We show that equilibrium SP is sender preferred if $U/c \in [\frac{p_d}{1-p_d}, \infty)$. In an equilibrium the agent sends a message that contains at least n outcomes on the equilibrium path. The ex ante expected utility is at most

$$\frac{1}{2}(U - nc) + \frac{1}{2}V_0\left(\frac{1}{2}\right) \tag{3.2}$$

because, if $s = 1$, then the agent gets his highest utility if he finds an outcome that can be used for persuasion in each experiment and if he can persuade the principal after n experiments. Note that we have $V_0(1/2) \leq 0$ in (3.2) due to Lemma 3.2.

We now show that the agent is strictly better off in equilibrium SP than in any other equilibrium with persuasion.

(A) If all messages $m^* \in M^*$ that are sent on the equilibrium path contain more than one outcome, then the equilibrium cannot be sender preferred as $n \geq 2$ and $V_0\left(\frac{1}{2}\right) \leq 0$ according to Lemma 3.2.

(B) There can be equilibria in which on the equilibrium path also messages are sent that contain only one outcome. For $n = 1$ the utility in (3.2) is weakly smaller than the utility in (3.1). We show now that equilibrium SP is also strictly preferred by the sender to these equilibria. We distinguish the following cases.

(a) Suppose an outcome of the first experiment that the agent runs is not an element of some m^* with $n = 1$. The agent needs to run a further experiment and $V_1(1/2) < (U - c)$. Thus, the agent prefers equilibrium SP.

(b) Suppose $y_1 = 0$ is an element of a message m^* with $n = 1$ that is sent on the equilibrium path, but $y_1 = 1$ is not an element of some message m^* with $n = 1$. We cannot have that the agent on the equilibrium path runs only one experiment, as by assumption $p_i^1 \geq 1 - p_i^0$ for all experiments i, which implies $prob\{s = 1 \mid y_1 = 0\} < 1/2 < p_d$. I.e., in such a hypothetical equilibrium the principal should not be persuaded upon the provision of a message that contains $y_1 = 0$ and no other outcome. Suppose the agent on the equilibrium path runs multiple experiments. In this case $V_1(1/2) < (U - c)$. Thus, the agent prefers equilibrium SP.

(c) Suppose $y_1 = 1$ is an element of a message m^* with $n = 1$ that is sent with positive probability on the equilibrium path. We now derive the maximum ex ante utility of the agent in such a potential equilibrium and properties of such a potential equilibrium. We then show that only equilibrium SP yields this utility.

Consider a potential equilibrium in which the precision of the first experiment is such that the principal is persuaded if the agent presents a positive outcome of this experiment. The maximum $V_1(1/2)$ that can potentially be achieved is $(U - c)$, since if $s = 1$ payoff $V_1(1/2) = (U - c)$ implies persuasion with probability 1 with the minimum number of experiments required for persuasion. $V_1(1/2) = (U - c)$ can be achieved in a potential equilibrium if and only if $p_1^1 = 1$. Thus, in such an equilibrium the posterior is $\overline{p} = 0$ if the first experimental outcome is adverse.

In such an equilibrium we cannot have that the agent continues running further experiments if he knows that $\overline{p} = 0$. In the hypothetical equilibrium either we have successful stopping if $y_1 = 1$ or the first experiment yields $y_1 = 0$ necessarily implying $\overline{p} = 0$. If the agent instead continues searching if $y_1 = 0$ knowing $\overline{p} = 0$, then he would not stop searching unsuccessfully in the following.[17] He would, thus, run experiments until he finds some $m^* \in M^*$ regardless

[17] Any following experiment does not change the posterior $\overline{p} = 0$, but at some later experiment the agent may have found some evidence that he can use for persuasion. Hence, the agent in the following cannot be worse off than after $y_1 = 0$.

of the state s. But then the principal should not be persuaded by such an m^* yielding a contradiction and, hence, the agent in equilibrium stops unsuccessfully if he knows that $\overline{p} = 0$.

As the agent stops searching unsuccessfully if $y_1 = 0$ in a hypothetical equilibrium with $V_1(1/2) = (U - c)$ the agent runs a single experiment on the equilibrium path. This determines the structure of $V_0(1/2)$ in such an equilibrium. As only one experiment is run on the equilibrium path, we have $V_0(1/2) = (1 - p_1^0)U - c$. The maximum $V_0(1/2)$ that can potentially be achieved in equilibrium is 0, as $V_0(1/2) \leq 0$ according to Lemma 3.2. Maximizing $V_0(1/2) = (1 - p_1^0)U - c$ with respect to p_1^0 subject to the constraint $V_0(1/2) \leq 0$ yields $p_1^0 = \frac{U-c}{U}$ which implies that $(1 - p_1^0)U - c = 0$. It follows that in the potential equilibrium which yields $V_1(1/2) = (U - c)$, we may also achieve $V_0(1/2) = 0$. It is, therefore, established that $p_1^1 = 1$ is the only possibility to have $V_1(1/2) = (U - c)$ and, given that $p_1^1 = 1$, $p_1^0 = \frac{U-c}{U}$ is the only precision in state 0 that potentially yields $V_0(1/2) = 0$ in equilibrium. It follows that $V_1(1/2) = (U - c)$ and $V_0(1/2) = 0$ can only be achieved if a single experiment is run on the equilibrium path with $(p_1^1, p_1^0) = (1, \frac{U-c}{U})$. An equilibrium in which the agent on the equilibrium path runs a single experiment and this experiment has precision $(p_1^1, p_1^0) = (1, \frac{U-c}{U})$ corresponds to equilibrium SP. Q.E.D.

Verification of Claim 2: In equilibrium we have $prob\{s = 1 \mid x = 1\} = prob\{s = 0 \mid x = 0\} = 1$: If there were an equilibrium with $prob\{s = 1 \mid x = 1\} < 1$, the principal would be persuaded by some message $m^{*\prime} \in M^*$ that exclusively contains outcomes from experiments with $p_i^0 < 1$. Then $m^{*\prime}$ can be found by chance even if $s = 0$ and the agent would run as many experiments as it takes to find some $m^* \in M^*$, as $U/c \to \infty$. He would find evidence rendering $m^{*\prime}$ feasible almost with certainty regardless of s. But then this message is worthless for the principal, contradicting that she is persuaded by $m^{*\prime}$. If there were an equilibrium with $prob\{s = 1 \mid x = 1\} = 1$ and $prob\{s = 0 \mid x = 0\} < 1$, then the persuasion probability would be smaller than 1/2. The agent could improve by running a single experiment which perfectly mirrors the state and yields persuasion probability 1/2. Without loss of generality, consider an equilibrium where the agent runs a single experiment on the equilibrium path that perfectly mirrors s. In this case, the persuasion probability is strictly lower than under public experimentation. Q.E.D.

Proof of Proposition 3.2: (i) The persuasion probability under private experimentation is $\frac{U+c}{2U}$, which is $\frac{U+c}{2U} = \frac{1}{2p_d}$ if $U/c = \frac{p_d}{1-p_d}$. As $\frac{U+c}{2U}$ decreases in U/c, the persuasion probability is smaller if $U/c > \frac{p_d}{1-p_d}$.

(ii) The decision quality is better under private experimentation, as $p_1^1 = 1$ under both schemes, but p_1^0 is higher under private experimentation. Q.E.D.

Proof of Proposition 3.3: (i) The higher U/c, the lower the probability to find favorable evidence conditional on $s = 0$ has to be in order to deter a further experiment. This is obtained by increasing p_1^0.

(ii) As $p_1^1 = 1$ does not change and p_1^0 increases according to (i), it follows that the persuasion probability decreases in U/c.

(iii) As $p_1^1 = 1$ does not change and p_1^0 increases according to (i), the statement directly follows. Q.E.D.

Chapter 4

Crisis and Credit Rating Agencies[1]

4.1 Introduction

Ratings and other quality certifications by third parties play an important role in today's economy. For instance, the volume of rated debt issues was over \$8,000 billion in 2006. Ratings are used by investors to guide their investment decisions. They are also crucial for financial regulation: Basel III includes ratings as one criterion for the calculation of the capital adequacy requirements for banks. So does the Solvency II Directive of the European Union, passed on March 11, 2014, which harmonizes insurance regulation in the European Union and is scheduled to come into effect on January 1, 2016.

However, ratings as a basis of regulation have been viewed controversially, especially after the financial crisis. The major concern is that the ratings used for regulation are given by rating agencies, which may have an incentive to distort ratings in order to maximize profit. As a reaction to this concern, Section 939A of the Dodd-Frank Act (effective since 2010) requires that all federal agencies "must remove any reference to or requirement of reliance on credit ratings".

The current article addresses the question of incentives to distort ratings by a profit maximizing rating agency under particular consideration of aggregate uncertainty. Aggregate uncertainty plays a major role in many markets. As an example, for subprime mortgages the question was not only how good the subprime mortgages were that one particular financial institution invested in. The question was whether subprime mortgages as a whole were a sufficiently safe investment.

To investigate the effect of aggregate uncertainty on incentives to distort, we consider a

[1]This chapter is joint work with Andras Niedermayer.

model in which all other possible incentives to distort are shut down. In particular, we consider a monopolistic rating agency that can credibly commit to a rating strategy in a one period model. This shuts down effects such as forum shopping, reneging on the ratings strategy, or reputational cycles.

Besides the rating agency there is a continuum of sellers selling bonds. There is a continuum of investors seeking to buy bonds. The mass of investors is larger than the mass of sellers, so that competition leads to prices being bid up to the expected value of a bond. The quality of a seller's bond is perfectly known to the seller, but unknown to investors. The rating agency has a technology to perfectly observe the seller's quality. Sellers can decide whether they want to be rated. The aggregate distribution of seller's types is initially unknown to all market participants, except for a common prior about the distribution of the aggregate states of the world. The states of the world differ by a different aggregate distribution of sellers' types. After sellers get rated, the aggregate state of the world is revealed to all market participants and investors buy the bonds. The price depends on the expected quality in a rating class for the realized aggregate state of the world.

We show that in accordance to the existing literature, a profit maximizing rating agency will choose a coarse binary rating: either investment grade or junk bonds. However, in sharp contrast to the existing literature, aggregate uncertainty leads to the cutoff not being at the first-best level. Whether the rating agency has an incentive to be too lenient (a negative cutoff) or too strict (a positive cutoff) is pinned down by three moments of the aggregate belief distribution. The aggregate belief distribution is defined as follows: Take for every state of the world the mean quality of bonds that would be bought in first-best. Market participants' belief distribution of these means is the aggregate belief distribution. The rating agency has more of an incentive to be too lenient if the distribution has a low mean, a high variance, and a low higher order skewness (defined as the sum of the third and higher moments). A low higher order skewness can be thought of as a left skewed distribution, i.e. with a high probability bonds have a mean quality above average, but the distribution has a fat tail at the bottom which implies that with a small probability bonds have a very low mean quality. The opposite result holds for a larger mean, lower variance, and a larger higher order skewness. These results can be interpreted as two opposite effects on the rating agency's incentive to distort ratings. One effect is pro-cyclical: they have an incentive to be too lenient before the outbreak of a crisis (interpreting this period as a period with a large variance and left skewness of aggregate uncertainty) and an incentive to be too strict after the outbreak of the crisis. The other effect is anti-cyclical: a higher mean in market beliefs about aggregate uncertainty (likely to occur before a crisis) gives the rating agency an incentive to be too strict and a lower mean (after a crisis) to be too lenient. While anecdotal evidence suggests that the pro-cyclical effect is stronger,[2] it is ultimately an empirical question, which effect dominates.

This sheds light on a disturbing aspect of using credit ratings for capital adequacy regulation: they may introduce pro-cyclicality into the system. Capital adequacy requirements based on ratings may be too lenient before and too strict after the crisis. Our theory can be seen to justify two possible policies to deal with this problem. One policy, as in Section 939A of the Frank-Dodd Act, is to remove any reference to or requirement of reliance on credit ratings from regulation. This approach has the advantage of having a clear unambiguous rule. However, this is also viewed controversially, since it may be too costly for smaller banks to replace external credit ratings with internal credit rating systems.[3] An alternative policy would be to use credit ratings, but take into account their cyclicality in regulation. In particular, if one believes that the pro-cyclical element dominates, capital adequacy requirements based on ratings should include anti-cyclical elements to counterbalance pro-cyclicality.

We provide two extensions of our main result. First, we outline an empirical strategy to determine whether the pro-cyclical or the counter-cyclical effect dominates. While an empirical analysis is beyond of the scope of this paper, we show how the moments of the distribution of aggregate uncertainty can be identified from the prices of financial derivatives.

Second, we extend the model to a setup with risk aversion. A model with risk aversion explains why there are multiple rating categories and not just one (i.e. investment grade, and possibly a second, speculative grade). The reason is that with risk aversion, investors value more precise information about the quality of an asset to reduce risk. We provide numerical examples to illustrate that a hybrid model of risk aversion and aggregate uncertainty preserves the key insights about the rating agency being too lenient or too strict, but additionally predicts multiple rating categories.

Our paper relates to a large literature on rating agencies, experts, and reputation. We differ from all papers mentioned below by having market participants' uncertainty about the

[2]In hindsight, observers of financial markets considered the ratings of agencies to have been too lenient before and too strict after the crisis.

[3]See, for example, http://www.americanbanker.com/bankthink/an-easy-fix-to-dodd-franks-credit-ratings-rule-1063396-1.html.

aggregate distribution of qualities as the driving force that determines the rating strategy.

If one were to remove aggregate uncertainty from our model, it would reduce to the model in Lizzeri (1999)'s seminal contribution on certification intermediaries. Lizzeri (1999) shows the by now well known result that certification intermediaries choose two categories (corresponding to investment grade and junk bonds) and set a cutoff at 0 which is the first-best level. (Note that this result can also be viewed as only one rating category being chosen – investment grade – and other assets not being rated.) Lizzeri (1999)'s work has been extended in a number of directions, including Doherty et al. (2012)'s work on risk-averse buyers. With risk-averse buyers, it can be optimal to have more than two categories.

Two papers allow for changes in the economic environment in a dynamic model. In Bolton et al. (2012) the rating agency trades off short term profits from consumers taking the rating at face value and long term reputational concerns. They assume that in a boom the fraction of naive consumers is high and, together with a low default risk, this gives the agency an incentive to inflate ratings during booms. Bar-Isaac and Shapiro (2013) investigate the quality of ratings when accuracy is costly for the agency. They combine reputational concerns with the change of economic fundamentals which affect, e.g., the costs for accuracy, possible profits and the default probability. They find that the rating quality is lower in booms than in recessions. Our analysis is complementary to these articles, since we show that a rating agency has an incentive to distort ratings even if all investors are rational and it is costless for the rating agency to assess the quality of the rating. Our results rely on the joint distribution of aggregate and idiosyncratic uncertainty.

In a wider sense, our paper also relates to the literature on experts and reputation. Reputation gives an incentive to report truthfully. Strausz (2005) shows that reputation leads to monopolization and that honest certification may require a price above that of a monopolist. Nevertheless, reputation is often not enough to ensure accurate information transmission (see Ottaviani and Sørensen, 2006; Bouvard and Levy, 2009; Mariano, 2008). Mathis et al. (2009) show that reputation and confidence cycles may exist, because the certifier likes to build up reputation so as to later inflate the grades and make larger profits.

The paper is structured as follows. Section 4.2 describes the model. Section 4.3 shows that it is optimal to rate according to a simple cutoff rule and Section 4.4 derives conditions under which this cutoff is positive or negative. Section 4.5 describes a stylized empirical identification strategy. Section 4.6 shows that with risk-averse investors several rating classes can be optimal but that the effects of aggregate uncertainty on the optimal cutoff remain. Section 4.7 concludes.

4.2 Model

There is one rating agency, a continuum of firms, and a continuum of possible investors. Each firm sells a good of quality t, where t is a random variable with support $[\underline{t}, \overline{t}]$ with $\underline{t} < 0 < \overline{t}$. The firm has private information about the quality. Investors are risk neutral and an investor's gross utility from buying the good is equal to the quality t.

There are N different states of the world. The probability of the world being in state i is ϵ_i. Having a two dimensional distribution (different states of the world, different distributions of qualities in each state of the world) adds a considerable amount of complexity. To still have a tractable model, we impose a restriction on this two dimensional distribution. We assume that there is a mass κ_i of sellers whose quality $\underline{t}$ is so low that one would never want to rate them (we will formalize this later on). There is a mass μ_i of sellers whose quality $\overline{t}$ is so high that one would always want to rate them. And then there is a mass λ_i of sellers with intermediate qualities $t \in (\underline{t}, \overline{t})$. We allow for arbitrary distributions of κ_i, λ_i, μ_i (with the only restrictions that the sum $\kappa_i + \lambda_i + \mu_i$ is constant and Assumption 1), but restrict the distribution conditional on being in $(\underline{t}, \overline{t})$ to be a distribution F which is the same for all states. We assume that F is continuously differentiable with density $f(t) > 0$ for all t in $(\underline{t}, \overline{t})$.

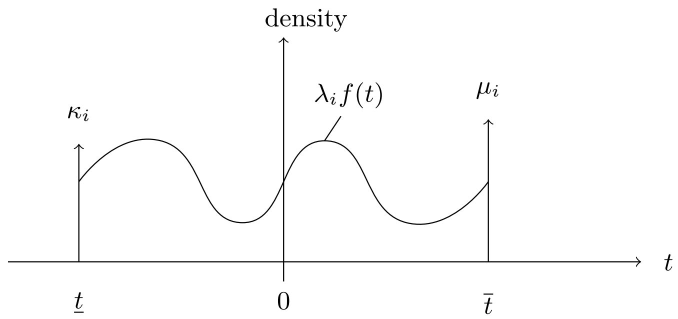

Figure 4.1: κ_i and μ_i are the mass points at $\underline{t}$ and $\overline{t}$ in state i. λ_i is the mass in state i that is allotted to the types $t \in (\underline{t}, \overline{t})$ with the distribution F.

Further, define the expected masses on $(\underline{t}, \overline{t})$ as $\tilde{\mu} := \sum_i \epsilon_i \mu_i$ and on $\overline{t}$ as $\tilde{\lambda} := \sum_i \epsilon_i \lambda_i$. Normalize $\tilde{\lambda}$ to 1. The probabilities ϵ_i and the distributions of quality are known to all players. We assume that $\underline{t}$ is sufficiently small:

Assumption 1.

$$\underline{t} < -\frac{\lambda_i \int_0^{\overline{t}} t dF(t) + \mu_i \overline{t}}{\kappa_i}, \qquad \forall i = 1, ..., N$$

Assumption 1 makes sure that we do not have to deal with the uninteresting corner solution in which the rating agency wants to rate all firms, including $\underline{t}$ firms.

A firm can choose to pay an upfront fee P to the rating agency in order to get rated before the state of the world becomes known to market participants. The agency rates firms that paid for a rating according to a precommitted rating strategy.[4]

The timing of moves is as follows:

- The agency sets the rating fee P and commits to a rating strategy s, $s(t) = r$, $s : \mathbb{R} \to \mathbb{R} \cup \{\emptyset\}$.
- Nature draws the state of the world i and quality t of each firm.
- The firms observe their own qualities, but not the state of the world, and decide whether to go to the agency to ask for a rating or not. This decision depends on the own type t, the strategy of the agency s and the price P.
- The agency observes the quality of the firms asking for ratings and gives ratings according to its strategy. The ratings are publicly observable. However, investors do not observe whether a firm went to the rating agency if the firm gets no rating ($\emptyset$).
- Observing the state of the world, the buyers decide how much to bid in a second price auction for a good. Since it is a second price auction, buyers bid their own expected valuation which depends on their belief about the expected quality given the information (s, P, r, i). Assuming that there are more investors than firms, investors will pay exactly the expected quality in equilibrium.

To solve the setup for equilibria we use Perfect Bayesian Equilibrium. We restrict the strategy of the firms to pure strategies and look at symmetric equilibria.

The profits of the agency in one state of the world is the rating fee P times the mass of firms asking for a rating. This mass depends on P and the rating strategy s. The agency is risk neutral and chooses s and P to maximize expected profits before knowing the state of the world.

The rating agency's rating strategy s partitions the set $[\underline{t}, \overline{t}]$ into M subsets, with each

[4]It does not matter in equilibrium whether the strategy is known at the beginning or not.

subset $m = 1, ..., M$ being the set of types $T_m = \{t|s(t) = r_m\}$ with M distinct r_m.[5] We will call these subsets rating classes in the following. Since in the end only the M distinguishable classes $\{T_m\}_{m=1}^M$ matter and not the labels $\{r_m\}_{m=1}^M$ attached to them, the following analysis will focus on $\{T_m\}$.

It is useful to define the expected quality in state i conditional on t being above a threshold $x > \underline{t}$ as

$$E_i(x) := \frac{\lambda_i \int_x^{\bar{t}} t dF + \mu_i \bar{t}}{\lambda_i \int_x^{\bar{t}} dF + \mu_i}.$$

A firm in $(\underline{t}, \bar{t})$ attaches probability $\hat{\epsilon}_i := \epsilon_i \lambda_i / \tilde{\lambda}$ to being in state i. Consequently, from a $(\underline{t}, \bar{t})$ firm's perspective, the expected quality above a threshold x over all states is

$$\tilde{E}(x) := \sum_i \hat{\epsilon}_i E_i(x).$$

In the following, we will assume that the virtual valuation function attached to $\tilde{E}(x)$ is monotone in x for $x \in (\underline{t}, \bar{t})$.

Assumption 2. $\tilde{E}(x) - \tilde{E}'(x) \frac{1-F(x)+\tilde{\mu}}{f(x)}$ is monotone in x for $x \in (\underline{t}, \bar{t})$.

This assumption basically ensures that the second-order condition is fulfilled whenever the first-order condition is fulfilled and it excludes the corner solution that it is optimal to only rate $\bar{t}$.

4.3 Optimality of Threshold Rating Strategy

In the following, we will show that it is optimal to rate all firms in an interval $[x, \bar{t}]$ in one rating class and not to give a rating to all firms with $t < x$. Formally, $s(t) = 1$ for all $t \geq x$ and $s(t) = \emptyset$ for all $t < x$.[6] We will show this in four steps. First, we show that it cannot be optimal to exclude type $\bar{t}$. Second, we show that the price of a rating is determined by firms with $t < \bar{t}$. Third, given that $\bar{t}$ is included, it is optimal to have only one rating class rather than multiple classes. Fourth, given that there is only one rating class, the set of types

[5] Technical speaking, there are $M + 1$ subsets because there can be types which do not receive any rating, $s(t) = \emptyset$. We will show in the following of this paper that it cannot be optimal to have more than two rating categories. Therefore, for the sake of notational simplicity, we do not consider an uncountable infinity of rating classes. To take into account the possibility of an uncountable infinity of rating classes, e.g. full disclosure, one could use the correspondence $T(r) = \{t|s(t) = r\}$ with $r \in \mathbb{R} \cup \{\emptyset\}$ instead of the sets $\{T_m\}_{m=1}^M$.

[6] This is equivalent to $s(t) = 1$ for all $t \geq x$ and $s(t) = 0$ for all $t < x$ because firms with $t < x$ are not rated in equilibrium.

belonging to this class has to be convex.

Lemma 4.1. *(i) It cannot be optimal that* $\underline{t} \in \cup_{m=1}^{M} T_m$. *(ii) It cannot be optimal that* $\bar{t} \notin \cup_{m=1}^{\tilde{M}} \tilde{T}_m$.

Part (i) of the Lemma holds by Assumption 1. The intuition for part (ii) of the Lemma is that $\bar{t}$ should be included in the rating because it increases the mass of rated firms as well as, due to its high type, other firms' willingness to pay for a rating.

Next, we state a lemma which will be useful throughout our analysis. The lemma states that if both firms with $t \in \left(\underline{t}, \bar{t}\right)$ and with $t = \bar{t}$ are in the same rating class, then firms with $t \in \left(\underline{t}, \bar{t}\right)$ have a lower willingness to pay for a rating then firms with $t = \bar{t}$.

Lemma 4.2. *Take an arbitrary rating class* T *that includes firms with* $t \in \left(\underline{t}, \bar{t}\right)$ *and* $t = \bar{t}$. *The willingness to pay for a rating is higher for* $\bar{t}$ *than for* $t \in (\underline{t}, \bar{t})$.

The reason is that firms update $\hat{\epsilon}_i$ differently and we show that firms with a type $\bar{t}$ assign a higher probability to states with higher expected quality than firms with $t \in (\underline{t}, \bar{t})$. Lemma 4.2 can be used to prove the next lemma, which states that if there are multiple rating classes and the highest type $\bar{t}$ is included, then it is better to merge all rating classes to one single class.

Lemma 4.3. $M = 1$ *with* $T_1 = \cup_{m=1}^{\tilde{M}} \tilde{T}_m$ *is better than* $\left\{\tilde{T}_m\right\}_{m=1}^{\tilde{M}}$ *with* $\tilde{M} > 1$ *if* $\bar{t} \in \cup_{m=1}^{\tilde{M}} \tilde{T}_m$.

Considering types that the agency intends to attract, the rating fee is always determined by the type with the lowest willingness to pay for a rating. Merging the rating class with a lowest willingness to pay with classes with a higher willingness to pay, the expected quality and thus, also the minimum willingness to pay increase.

The next lemma states that all firms above a threshold are rated which means that no types in between are excluded.

Lemma 4.4. *If* $M = 1$ *and* $\bar{t} \in T_1$, *then* T_1 *has to be convex.*

If the set is not convex, there is at least one unrated hole in the middle and the agency can rate firms in the hole instead of rating some other types below with the same mass. This increases the expected type in every state and, therefore, also the rating fee the agency can charge from the firms.

Lemmas 4.1, 4.3, and 4.4 together lead to the following proposition.

Proposition 4.1. *It is optimal to choose* $M = 1$ *with* $T_1 = [x, \bar{t}]$ *for some* x.

Proposition 4.1 shows that the best equilibrium for the rating agency is such that the agency offers the following ratings strategy:

$$s(t) = \begin{cases} 1 & \text{if } t \geq x, \\ \emptyset & \text{otherwise,} \end{cases}$$

that is, all firms above some cutoff x get a positive rating. Subsequently, all firms with $t \in [x, \bar{t}]$ get rated and investors pay the expected quality over $[x, \bar{t}]$.

As usual in such models, there is a multiplicity of equilibria in the subgame following the ratings agency's announcement of its price P and rating strategy s. For example, there is the trivial equilibrium in which no firm applies for a rating and investors have the off-equilibrium belief that firms that do get a rating are of the worst possible rated quality x. Since x is less than the price of a rating P, it is a best response for firms to stay unrated.

The usual arguments for selecting the equilibrium we described apply: The rating agency has a first-mover advantage, hence, it is reasonable that the equilibrium most favorable to the rating agency will be selected. Further, by a small perturbation of its strategy, the rating agency can get rid of undesired equilibria. For example, if no firm gets a rating, the agency might incentivize the first firms who apply for a rating in order to jump-start the market.[7]

4.4 Optimal Threshold

By Proposition 4.1 we can restrict our attention to threshold rules which consist of all types above a cutoff x being pooled in one class and all types below not being rated. If there were only one state of the world, the optimal threshold would be $x = 0$. To see this, take a model with only one state of the world, e.g. by setting $\mu_i = \tilde{\mu}$ and $\lambda_i = \tilde{\lambda} = 1$ for all i. Then the agency's profit is

$$\Pi = (1 - F(x) + \tilde{\mu}) \frac{\int_x^{\bar{t}} t dF(t) + \tilde{\mu}\bar{t}}{1 - F(x) + \tilde{\mu}} = \int_x^{\bar{t}} t dF(t) + \tilde{\mu}\bar{t}.$$

[7] A simple, albeit extreme example is the following: As long as not all firms with a quality $t \in [x, \bar{t}]$ enter, firms get their rating fees refunded and get an additional small compensation. This makes sure that any equilibrium in which not all firms in $[x, \bar{t}]$ get rated breaks down, so that the refund never has to be paid in equilibrium.

which is equal to welfare. The first derivative is $\frac{\partial \Pi}{\partial x} = -xf(x)$, which is equal 0 if $x = 0$. Therefore, the optimal threshold for the agency is $x = 0$.[8] This special case of our model corresponds to Lizzeri (1999)'s results.

If there are N states of the world, the rating agency's profit is

$$\begin{aligned}\Pi(x) :=& \left(\sum_{i=1}^{N}(\lambda_i(1-F(x))+\mu_i)\epsilon_i\right)\left(\sum_{j=1}^{N}E_j(x)\hat{\epsilon}_j\right)\\ =&(1-F(x)+\tilde{\mu})\tilde{E}(x)\end{aligned}$$

where $\tilde{E}(x)$ is the expected value of a rating from a firm's perspective which assigns the probabilities $\hat{\epsilon}_i$ to different states.

The welfare with N states of the world is

$$\begin{aligned}W(x) :=& \sum_{i=1}^{N}\left(\lambda_i\int_x^{\bar{t}} tdF(t)+\mu_i\bar{t}\right)\epsilon_i\\ =&\sum_{i=1}^{N}E_i(x)(\lambda_i(1-F(x))+\mu_i)\epsilon_i.\end{aligned}$$

Define the expected type on $[x, \bar{t})$ as

$$E_0(x) := \frac{\int_x^{\bar{t}} tdF(t)}{1-F(x)}.$$

Rearrange the expression for the welfare to

$$\begin{aligned}W(x) =& \sum_i(\lambda_i(1-F(x))E_0(x)+\mu_i\bar{t})\epsilon_i\\ =&(1-F(x)+\tilde{\mu})\hat{E}(x)\end{aligned}$$

with

$$\hat{E}(x) := \frac{(1-F(x))E_0(x)+\tilde{\mu}\bar{t}}{1-F(x)+\tilde{\mu}},$$

which can also be written as

$$\hat{E}(x) = \frac{\sum_i \epsilon_i(\lambda_i(1-F(x))+\mu_i)E_i(x)}{1-F(x)+\tilde{\mu}}.$$

[8]It is easy to check that the second-order condition is also satisfied at $x = 0$.

$\hat{E}(x)$ is the expected value of a rating from a welfare perspective which takes into account that the quantity of firms being rated $(\lambda_i(1-F(x))+\mu_i)$ is different in every state. In the following, we will drop the argument x in $E_i(x)$, $E_0(x)$, $\tilde{E}(x)$, $\hat{E}(x)$ when it is unambiguous in order to simplify notation. $\hat{E}$ and $\tilde{E}$ compare in the following way.

Lemma 4.5. *The value of a rating is larger from a welfare then from a firm's perspective;* $\hat{E} \geq \tilde{E}$ *for all* x.

This implies that $W(x) \geq \Pi(x)$. For non-degenerate distributions of the state of the world, the inequality is strict and the rating agency cannot extract the whole surplus, $W(x) > \Pi(x)$.[9]

The derivative of the profit with respect to the cutoff is

$$\begin{aligned}\frac{\partial \Pi}{\partial x} &= (1-F(x)+\tilde{\mu})\sum_i \hat{\epsilon}_i \frac{\partial E_i}{\partial x} - f(x)\tilde{E}(x) \\ &= -f(x)\left[\underbrace{\tilde{E}(x)}_{marginal\ effect} - \underbrace{\frac{1-F(x)+\tilde{\mu}}{f(x)}\frac{\partial \tilde{E}}{\partial x}}_{inframarginal\ effect}\right].\end{aligned} \tag{4.1}$$

and we will show later that the first order condition is sufficient for profit maximization. Thus, the profit maximizing cutoff is given by $\Pi'(x) = 0$. Changing the cutoff has two opposite effects on the agency's profit; increasing the cutoff decreases the mass of firms asking to be rated (marginal effect) but it also increases the expected quality of firms being rated and by this it increases a firm's willingness to pay for being rated (inframarginal effect).

We call the expression in the squared brackets in (4.1) the virtual valuation function for $\tilde{E}$.[10] By Assumption 2 it is monotone and, thus, the first order condition is sufficient to find an optimum.[11] This also implies that there is a unique solution of the first order condition.

We are interested in comparing the profit maximizing cutoff with the welfare maximizing cutoff. Thus, we also have to determine the socially optimal cutoff. The derivative of welfare

[9]Even if $\hat{\epsilon}_i = \epsilon_i$ for all i, the inequality is strict for non-degenerated distributions. Besides by the updating of $\hat{\epsilon}_i$, the difference between $\hat{E}$ and $\tilde{E}$ is caused by the different mass of firms being rated in different states of the world.

[10]We can rewrite the virtual valuation in terms of E_i as $\sum_i \epsilon_i\lambda_i\left(E_i - E_i'\frac{1-F+\tilde{\mu}}{f}\right)$.

[11]The second order condition follows directly from Assumption 2. That we do not have a corner solution at $x = \bar{t}$ can be seen by observing that the profit $\Pi(x)$ is continuous at $x = \bar{t}$ and $\lim_{x\to\bar{t}}\Pi'(x) < 0$. Assumption 1 implies that there is no corner solution at $x = \underline{t}$ (see proof of Lemma 4.1).

with respect to the threshold is

$$\frac{\partial W}{\partial x} = -f(x)\left(\underbrace{\hat{E}(x)}_{marginal\ effect} - \underbrace{\frac{1-F(x)+\tilde{\mu}}{f(x)}\frac{\partial \hat{E}}{\partial x}}_{inframarginal\ effect}\right). \tag{4.2}$$

or written in a simpler way

$$\frac{\partial W}{\partial x} = -\sum_i \hat{\epsilon}_i \lambda_i x f(x) = -xf(x)$$

which is the same as for one state of the world. The derivative is 0 if $x = 0$ and thus, the welfare maximizing cutoff is at 0.

To derive the difference between the profit of an agency and the welfare, we write the profit as

$$\begin{aligned}\Pi(x) =& \left(\sum_i (\lambda_i(1-F(x)) + \mu_i)\epsilon_i\right)\left(\sum_j E_j(x)\hat{\epsilon}_j\right)\\ =& \sum_j \left(E_j(x)\hat{\epsilon}_j\left(\sum_i (\lambda_i(1-F(x))+\mu_i)\epsilon_i\right)\right)\\ =& \sum_j E_j(x)(1-F(x)+\mu_j)\hat{\epsilon}_j - \sum_j E_j(x)\hat{\epsilon}_j\left(\mu_j - \sum_i \mu_i\epsilon_i\right)\\ =& W(x) + L(x)\end{aligned}$$

where $L(x) := -\sum_j E_j\hat{\epsilon}_j(\mu_j - E[\mu])$ is the non-extractable part of the surplus ("loss" compared to extracting total surplus). Remember that $W'(0) = 0$ which implies that $L'(0) = \Pi'(0)$. Thus, the incentive for the agency to distort the rating compared to the welfare maximizing rating is given by the sign of $L'(0)$. The optimal cutoff is positive if $L'(0) > 0$ and it is negative if $L'(0) < 0$.

Proposition 4.2. *The derivative $L'(0)$ is given by*

$$L'(0) = \frac{f(0)\tilde{E}}{\bar{t}-\hat{E}}\left(\hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}\right). \tag{4.3}$$

Since the expression before the parenthesis is always positive, the sign of $L'(0)$ and, therefore, the sign of the profit maximizing cutoff depends on the sign of $\left(\hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}\right)$.

$\hat{E}-\tilde{E}$ is positive and can be interpreted as the difference of the expected value of a rating from a social and a firm's perspective. An intuition for $\frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}$ is that it is the variance divided by the mean of the posterior distribution of E_i and it reflects the uncertainty about the state of the world: if this uncertainty is sufficiently large, the cutoff is negative. The reason for this is that firms care less about the effect of the cutoff x on the expected quality of a rated firm if the expected quality is to a large extent driven by uncertainty about the state of the world. Thus, the sign of $L'(0)$ is determined by the difference of the expected value of a rating and the ratio of variance to mean of the posterior distribution of E_i.

While the above expression for $L'(0)$ provides some insights on the determinants of the optimal cutoff, it is difficult to use it for comparative statics, since a change of the mean and variance of E_i will also change $\hat{E}$. Therefore, in the following, we will express $L'(0)$ in terms of the moments of the posterior distribution of E_i. From $\mu_i = \lambda_i(1-F(x))\frac{E_i - E_0}{\bar{t}-E_i}$, we can write

$$\begin{aligned}\tilde{\mu} &= \sum_i \epsilon_i \lambda_i (1-F(x)) \frac{E_i - E_0}{\bar{t}-E_i} \\ &= (1-F(x)) \sum_i \hat{\epsilon}_i \frac{E_i - \bar{t} + \bar{t} - E_0}{\bar{t}-E_i} \\ &= (1-F(x)) \left(-1 + (\bar{t}-E_0) \sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-E_i} \right). \end{aligned} \tag{4.4}$$

Plugging (4.4) into the definition of $\hat{E}$ we get

$$\begin{aligned}\hat{E} =& \frac{(1-F(x))E_0 + \tilde{\mu}\bar{t}}{1-F(x)+\tilde{\mu}} \\ =& \frac{E_0 + \left(-1 + (\bar{t}-E_0)\sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-E_i}\right)\bar{t}}{1 + \left(-1 + (\bar{t}-E_0)\sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-E_i}\right)} \\ =& \bar{t} - \frac{1}{\sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-E_i}}. \end{aligned}$$

Define the scaled value of a rating as $e_i := \frac{E_i}{\bar{t}}$. Then $\sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-E_i} = \sum_i \hat{\epsilon}_i \frac{1}{\bar{t}-e_i\bar{t}}$. The kth derivative of $\frac{1}{\bar{t}(1-e_i)}$ with respect to e_i is

$$\frac{\partial^k}{\partial e_i^k}\left[\frac{1}{\bar{t}(1-e_i)}\right] = k!\frac{1}{\bar{t}(1-e_i)^{k+1}}.$$

Using these derivatives one can construct a Taylor series of $\frac{1}{\bar{t}(1-e_i)}$ with respect to e_i around $e_i = 0$. This yields

$$\begin{aligned}\frac{1}{\bar{t}(1-e_i)} &= \sum_{k=0}^{\infty} \frac{e_i^k}{k!} \frac{\partial^k}{\partial e_i^k} \left[\frac{1}{\bar{t}(1-e_i)} \right] \Bigg|_{e_i=0} \\ &= \sum_{k=0}^{\infty} \frac{e_i^k}{\bar{t}}.\end{aligned}$$

Taking expectations over the state of the world yields

$$\sum_i \hat{\epsilon}_i \left[\frac{1}{\bar{t}(1-e_i)} \right] = \frac{1}{\bar{t}} \left(1 + m_1 + m_2 + \sum_{k=3}^{\infty} m_k \right),$$

where $m_k := \sum_i \hat{\epsilon}_i e_i^k$ is the kth moment of the posterior distribution of e_i. This implies that we can write

$$\hat{E} = \bar{t} - \frac{\bar{t}}{1 + m_1 + m_2 + \sum_{k=3}^{\infty} m_k}. \tag{4.5}$$

Define $m_{3+} := \sum_{k=3}^{\infty} m_k$. Observe that (4.3) simplifies to

$$L'(0) = \frac{f\tilde{E}}{\bar{t} - \hat{E}} \left(\hat{E} - \frac{\sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}} \right). \tag{4.6}$$

Plugging (4.5), $\tilde{E} = m_1 \bar{t}$ and $\sum_i \hat{\epsilon}_i E_i^2 = \bar{t}^2 m_2$ into (4.6) yields

$$\begin{aligned}L'(0) =& \frac{f m_1 \bar{t}}{\frac{\bar{t}}{1+m_1+m_2+m_{3+}}} \left(\bar{t} - \frac{\bar{t}}{1 + m_1 + m_2 + m_{3+}} - \frac{\bar{t}^2 m_2}{m_1 \bar{t}} \right) \\ =& f m_1 \bar{t} \left(\sum_{k=0}^{\infty} m_k \right) \underbrace{\left[1 - \frac{1}{1 + m_1 + m_2 + m_{3+}} - \frac{m_2}{m_1} \right]}_{=:S}.\end{aligned} \tag{4.7}$$

The sign of $L'(0)$ is given by the sign of S. Note that S only depends on the moments of e_i, more precisely, it depends only on the mean m_1, the second moment m_2 and the sum of all higher moments m_{3+}. For example, let us start with $L'(0) < 0$. If we keep mean and second moment constant and increase the sum of higher moments, S increases and $L'(0)$ can switch sign from negative to positive. This means that we change the optimal cutoff from a negative to a positive one by changing the higher moments of the distribution of e_i.

We can calculate the threshold $\overline{m}_{3+}$ for which $L'(0)$ is 0. Set

$$1 - \frac{1}{1 + m_1 + m_2 + \overline{m}_{3+}} - \frac{m_2}{m_1} = 0$$

which leads to

$$\overline{m}_{3+} = \frac{m_2^2 + m_2 - m_1^2}{m_1 - m_2}.$$

Observe that $\overline{m}_{3+}$ is always positive because $m_1 > m_2$ and $m_2 - m_1^2$ is the variance of e_i. This implies that for $m_{3+} < \overline{m}_{3+}$ the expression S is negative and thus $L'(0) = \Pi'(0) < 0$.

Proposition 4.3. *The optimal cutoff for the rating agency is negative if $m_{3+} < \overline{m}_{3+}$ and positive if $m_{3+} > \overline{m}_{3+}$.*

We also derive thresholds for m_1 and m_2. First, observe that S is increasing in m_1 and decreasing in m_2 given that $m_1, m_2, m_{3+} > 0$. Second, by setting the expression in square brackets to zero and solving for m_1 and m_2, respectively, one gets thresholds for m_1 and m_2 that determine whether the cutoff of the rating agency is positive or negative. The thresholds are stated in the following two Propositions.

Proposition 4.4. *The optimal cutoff for the rating agency is negative if $m_1 < \overline{m}_1$ and positive if $m_1 > \overline{m}_1$, where*

$$\overline{m}_1 := \frac{1}{2}\left(-m_{3+} + \sqrt{4m_2 + (2m_2 + m_{3+})^2}\right)$$

The optimal cutoff for the rating agency is negative if $m_2 > \underline{m}_2$ and positive if $m_2 < \underline{m}_2$, where

$$\underline{m}_2 := \frac{1}{2}\left(-1 - m_{3+} + \sqrt{2m_{3+} + 1 + (2m_1 + m_{3+})^2 + 2m_{3+} + 1}\right)$$

Both thresholds, $\overline{m}_1$ and $\underline{m}_2$, are positive given that $m_1, m_2, m_{3+} > 0$.

Propositions 4.3, 4.4, and 4.4 have a striking implication: the rating agency has more of an incentive to be too lenient if the distribution of aggregate uncertainty is more left skewed (in the sense of a smaller higher order skewness or low m_{3+}), the mean is smaller, or the variance is larger. Left skewness and a high variance can be reasonably considered as being associated with a period preceding the beginning of a crisis. For moments that can be reasonably associated with a period shortly after a crisis (right skewness, low variance), incentive of the rating agency move in the opposite direction: the rating agency has an increasing incentive to be too strict. This gives the rating agency an incentive to rate pro-cyclically: excessively lenient ratings expand investments during booms, excessively restrictive ratings restrict investments during recessions. Observe that the mean of aggregate uncertainty has a counter cyclical effect; a small expected average, which can be associated with a period shortly after a crisis, gives the rating agency an incentive to be too lenient. The opposite holds for a high expected average.

4.4.1 Example of Beta Distributions

It is illustrative to parametrize the posterior distribution of E_i as a Beta distribution with support $[E_0, \bar{t}]$, i.e. E_i having a density $h(y) \propto y^{\alpha-1}(1-y)^{\beta-1}$, where $y = (E_i - E_0)/(\bar{t} - E_0)$. The distribution of $E_i/\bar{t}$ is determined by the three parameters α, β, and $e_0 := E_0/\bar{t}$. (The upper bound of the support of $E_i/\bar{t}$ is 1.) These three parameters determine m_1, m_2, and m_{3+} the following way:

$$\begin{aligned}
m_1 &= \frac{\alpha + \beta e_0}{\alpha + \beta}, \\
m_2 &= \frac{(1-e_0)^2 \alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)} + m_1^2 \\
m_{3+} &= \frac{\alpha + \beta - 1}{(1-e_0)(\beta - 1)} - 1 - m_1 - m_2
\end{aligned}$$

It can be shown that this is a one-to-one mapping from (α, β, e_0) to (m_1, m_2, m_{3+}).[12] One can use this one-to-one mapping for comparative statics with respect to say m_{3+} while keeping m_1 and m_2 constant. Figure 4.2 shows a Beta distribution with $\alpha = 3$, $\beta = 5$ and $e_0 = 0.1$

[12] The mapping in the opposite direction can be derived in closed form, but the resulting expressions are rather long and uninformative and therefore omitted. m_1 and m_2 are the well-known first two moments of the Beta distribution. m_{3+} can be derived by observing that $E[(1-y)^{-1}(1-e_0)^{-1}] = E[\sum_{k=0}^{\infty}(e_0 + (1-e_0)y)^k] = E[\sum_{k=0}^{\infty} e^k] = 1 + m_1 + m_2 + m_{3+}$, where $e = E_i/\bar{t} = e_0 + (1-e_0)y$. For a Beta distribution with density $h(y) = y^{\alpha-1}(1-y)^{\beta-1}/B(\alpha,\beta)$ the expected value is

$$E\left[\frac{1}{1-y}\right] = \int_0^1 \frac{y^{\alpha-1}(1-y)^{\beta-2}}{B(\alpha,\beta)} dy = \frac{B(\alpha, \beta-1)}{B(\alpha,\beta)} = \frac{\alpha+\beta-1}{\beta-1},$$

where the last equality follows from the relation of the Beta to the Gamma function

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

and the property $\Gamma(x+1) = x\Gamma(x)$ of the Gamma function which imply

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta-1)(\beta-1)}{\Gamma(\alpha+\beta-1)(\alpha+\beta-1)} = \frac{\beta-1}{\alpha+\beta-1} B(\alpha, \beta-1).$$

Putting this together yields the expression for m_{3+}.

(dashed line). For this distribution, $L'(0) = 0$, i.e. the rating agencies sets the cutoff at exactly the socially optimal level $x = 0$. For the dotted line, m_1 and m_2 are kept constant and m_{3+} is reduced by 0.01. The dotted line has a fatter lower tail which means that it has a higher mass at the bottom of the distribution. The mean and variance remain the same, but if a crisis hits, it is more likely to be severe. For the dotted distribution $L'(0) < 0$ and hence the cutoff is negative, $x < 0$, which means that the rating criteria are too loose compared to the socially optimal ones. For the solid line, m_{3+} is increased by 0.01 while keeping m_1 and m_2 constant. For this distribution $L'(0) > 0$ and hence $x > 0$, that is, the rating is too strict compared to the socially optimal one.

Figures 4.3, 4.4, and 4.5 illustrate the change of $L'(0)$ as m_{3+}, m_1, and m_2 are changed, respectively, while keeping the other parameters constant. The optimal cutoff for example can switch from a negative to a positive cutoff if the mean or the higher order skewness increase or if the variance decreases. For all values of m_1, m_2, and m_{3+}, the parameters α, β, and e_0 are in permissible ranges.[13]

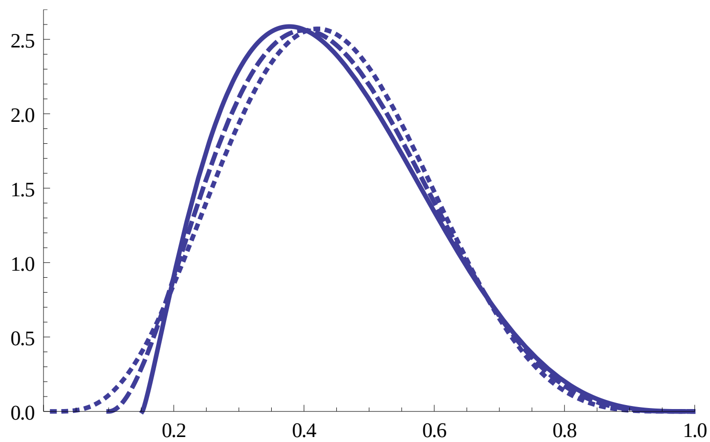

Figure 4.2: Density of $E_i/\bar{t}$ for $\alpha = 3$, $\beta = 5$, $e_0 = 0.1$ (dashed line). For the dotted line, m_{3+} is reduced by 0.01, for the solid line, m_{3+} is increased by 0.01, while m_1 and m_2 are kept constant. (The corresponding parameters are $\alpha = 4.4322$, $\beta = 5.8781$, $e_0 = 0.013363$ for the dotted and $\alpha = 2.23$, $\beta = 4.38985$, $e_0 = 0.151755$ for the solid distribution.)

4.5 Empirical Implications

Our model shows how the rating agency's incentive to be too lenient or too strict depends on the moments of aggregate uncertainty. Since these moments cannot be observed directly, one may wonder about the empirical content of our model.

[13]The permissible ranges are $\alpha > 0$, $\beta > 0$, and $e_0 \in (0, 1)$.

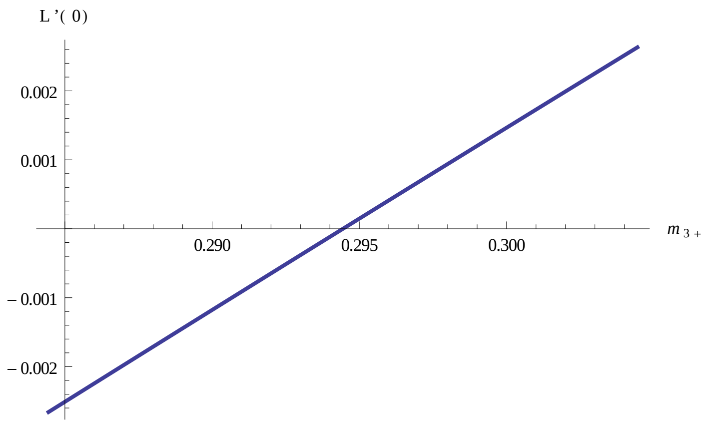

Figure 4.3: Values of $L'(0)$ as m_{3+} is changed and m_1 and m_2 are kept constant. Starting point is $\alpha = 3$, $\beta = 5$, $e_0 = 0.1$ (which corresponds to $m_1 = 0.4375$, $m_2 = 0.2125$, and $m_{3+} = 0.294444$) for which $L'(0) = 0$. Further parameters are normalized to $\bar{t} = 1$ and $f(0) = 1$.

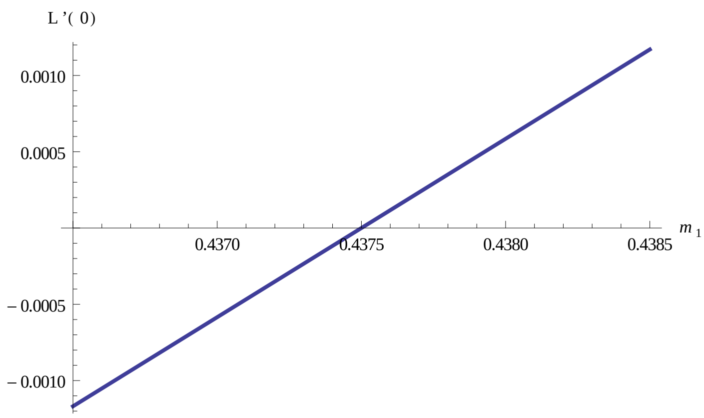

Figure 4.4: Values of $L'(0)$ as m_1 is changed and m_2 and m_{3+} are kept constant. Starting point is $\alpha = 3$, $\beta = 5$, $e_0 = 0.1$ (which corresponds to $m_1 = 0.4375$, $m_2 = 0.2125$, and $m_{3+} = 0.294444$) for which $L'(0) = 0$. Further parameters are normalized to $\bar{t} = 1$ and $f(0) = 1$.

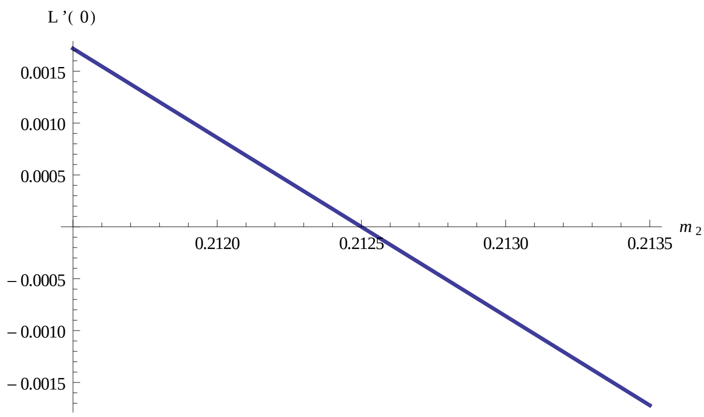

Figure 4.5: Values of $L'(0)$ as m_2 is changed and m_1 and m_{3+} are kept constant. Starting point is $\alpha = 3$, $\beta = 5$, $e_0 = 0.1$ (which corresponds to $m_1 = 0.4375$, $m_2 = 0.2125$, and $m_{3+} = 0.294444$) for which $L'(0) = 0$. Further parameters are normalized to $\bar{t} = 1$ and $f(0) = 1$.

First, it should be noted that an empirical estimate of the distribution of aggregate uncertainty is non-trivial, especially if the main concern is about the distribution of aggregate uncertainty shortly before a crisis. The reason is that only few crises occur, so it is difficult to have larger amounts of data.

However, an empirical estimate of market participants' beliefs about the distribution of aggregate uncertainty can be obtained. We illustrate the basic idea of how to estimate these moments in a strongly stylized setup containing the core idea of the empirical strategy.

Consider the following stylized setup. There is an index for the bonds being sold by the firms in the market. Further, there is a market for financial derivatives based on this index. As an example, one can think of an index on securitized assets backed by subprime mortgages. Call options on the index can be bought in the first period of the model, before aggregate uncertainty is realized. The options expire in the second period after aggregate uncertainty has realized. Time is discrete and the options are European options.[14] Further, aggregate uncertainty is such that the mid-quality firms' beliefs are the same as the general market beliefs, formally, $\hat{\epsilon}_i = \epsilon_i$ for all i.[15] Suppose that the cut-off of the agency is close to 0 ($x \approx 0$), so that the value of the index $E_i(x)$ is well approximated by $E_i(0)$.

Further, assume that there exist a call option with strike price $y_i = E_i$ for each state of the world i. Without loss of generality, order the states of the world increasingly, i.e. $E_j > E_i$ if $j > i$. The second-period value of a call option with strike price y_j in state i is $E_i - y_j$ if $E_i > y_j$ and 0 if $E_i \leq y_j$. Denote the first-period price of option j with strike price y_j as O_j. O_j is given by the market's expected value of the second period value (ignoring discounting):

$$O_j = \sum_{i=1}^{N} \epsilon_i \max\{E_i - y_j, 0\} = \sum_{i=j+1}^{N} \epsilon_i (E_i - E_j) \tag{4.8}$$

where the second equality follows from $y_j = E_j$. (For $i = N$, $O_j = 0$.)

The next proposition shows that given a set of call options, the information on their strike prices y_j and first-period prices O_j identifies the market's beliefs about the distribution of aggregate uncertainty; it identifies the probability ϵ_i for the expected quality E_i.

Proposition 4.5. *Given strike prices and first period prices* $\{(y_j, O_j)\}_{j=1}^{N}$*, the probability*

[14] In a discrete two-period model, it does not matter whether the option is European or American. In a continuous time model, calculations for American options are somewhat more complex, but standard and well known in the literature.

[15] A sufficient condition is that $\lambda_i = \tilde{\lambda}$ for all i, that is, aggregate uncertainty enters through changes of κ_i and μ_i for different states of the world i.

mass function of the distribution of aggregate uncertainty is given by

$$\epsilon_j = \frac{O_j - O_{j+1}}{y_{j+1} - y_j} - \frac{O_{j-1} - O_j}{y_j - y_{j-1}}$$

for $1 < j < N$ *and*

$$\epsilon_N = \frac{O_{N-1}}{y_N - y_{N-1}},$$
$$\epsilon_1 = 1 - \sum_{i=2}^{N} \epsilon_i.$$

A similar result can be obtained for a continuous distribution of E_i. For the continuous distribution version, drop the index in E_i and denote the distribution of E as G. Assume that prices $O(y)$ for call options with a continuum of strike prices $y \in [\underline{t}, \overline{t}]$ are observed. Then $O(y)$ is given by

$$O(y) = \int_{\underline{t}}^{\overline{t}} \max\{E - y, 0\} dG(E) = \int_{y}^{\overline{t}} (E - y) dG(E).$$

The first derivative is

$$O'(y) = \int_{y}^{\overline{t}} (-1) dG(c) - [(y - y) g(y)] = -(1 - G(y)),$$

and the second

$$O''(y) = g(y).$$

This is analogous to the discrete distribution result and the distribution G is non-parametrically identifiable given data on call option prices.

In practice, one expects to observe less options than there are states of the world, so parametric assumptions are required to be able to estimate the distribution of E.

In the following, we make the parametric assumption that the distribution G is a polynomial with lower bound of support E_0 and upper bound $\overline{t}$. As an example, consider a cubic function

$$G(E) = a_1 + 2a_2 E + 3a_3 E^2 + 4a_4 E^3.$$

The price of a call option will also be a polynomial function of the strike price y, since

$$O(y) = \int_y^{\bar{t}} (1 - G(E))dE = a_0 + a_1 y + a_2 y^2 + a_3 y^3 + a_4 y^4,$$

where

$$a_0 = -\sum_{i=1}^{4} a_i \bar{t}^i.$$

Suppose we observe data for five call options with strike prices $\{y_j\}_{j=1}^5$ and option prices $\{O(y_j)\}_{j=1}^5$. In this case, the parameters $\{a_i\}_{i=0}^4$ are given by the linear equation system

$$O(y_j) = a_0 + a_1 y_j + a_2 y_j^2 + a_3 y_j^3 + a_4 y_j^4, \qquad j = 1, ..., 5. \tag{4.9}$$

As long as the matrix $\left[y_j^i\right]_{j=1,...,5; i=0,...,4}$ is non-singular, the equation system (4.9) yields a unique solution for the variables $\{a_i\}_{i=0}^4$. Note that E_0 and $\bar{t}$ are uniquely pinned down by the parameters $\{a_i\}_{i=0}^4$ and by the equations $G(E_0) = 0$ and $G(\bar{t}) = 0$.[16]

Given the distribution G of E, we can obtain the distribution of $e = E/\bar{t}$ and the moments m_1, m_2, m_{3+} of e. This in turn yields

$$S = 1 - \frac{1}{1 + m_1 + m_2 + m_{3+}} - \frac{m_2}{m_1}$$

from expression (4.7) and determines the sign of the marginal profit $\Pi'(0)$ at $x = 0$. Table 4.1 provides examples of observed prices of call options and corresponding estimated parameters, moments, and S. For the first set of observations (first line), the rating agency has an incentive to choose the cutoff at the first best level $x = 0$. For the second line and third line, the agency has an incentive to choose a negative and a positive cutoff, respectively.

observed prices					estimated parameters				moments			S
O_1	O_2	O_3	O_4	O_5	a_1	a_2	a_3	a_4	m_1	m_2	m_{3+}	
11	8.0	5.7	3.9	2.6	-0.37	0.0096	-0.000030	3.4×10^{-8}	0.35	0.15	0.30	0.0
5.9	4.1	2.7	1.8	1.1	0.0	0.0077	-0.000026	3.4×10^{-8}	0.24	0.095	0.12	-0.073
14	10	7.2	4.9	3.2	-0.89	0.014	-0.000043	5.1×10^{-8}	0.41	0.20	0.37	0.019

Table 4.1: Example of parameter estimates for data on call option prices $O_j = O(y_j)$ for strike prices $(y_1, y_2, y_3, y_4, y_5) = (80, 90, 100, 110, 120)$.

We have illustrated the basic idea behind an empirical strategy to estimate the moments of

[16]While the $G(E)$ has multiple roots due to G being a polynomial, the solution of $G(E_0) = 0$ is unique nonetheless. This is because of the constraints $G'(E) > 0$ for $E \in [E_0, \bar{t}]$ and $y_j \in [E_0, \bar{t}]$ for all j. By the same reasoning, there is a unique solution of $G(\bar{t}) = 1$.

aggregate uncertainty. To practically apply this strategy, several additional steps are required, which are beyond the scope of this article. First, one needs to construct synthetic call options for the index of the bonds being rated. Second, the pricing of options in a multi-period environment is more complicated than the simple two-period setup we used for illustrative purposes. These problems are far from trivial, but well studied in Finance, see e.g. Hull (2009). Additionally, one could use a different parametrization for G instead of the polynomial parametrization or, if sufficiently many observations are available, one could possibly even use a non-parametric estimate of the function $O(y)$ given the observations $\{y_j, O(y_j)\}_j$. Further, one would also want to estimate the confidence interval for S.

4.6 Risk-aversion

In the main part of this paper we have assumed that investors are risk neutral and we have shown that it is optimal for the agency to pool all types above a cutoff in one rating class. Doherty et al. (2012) extend the model of Lizzeri (1999) by allowing investors to be risk-averse and they show that, if the level of risk aversion is sufficiently high, the rating agency rates types above a cutoff in several rating classes.

First, following the paper of Doherty et al. (2012) we provide a simplified hybrid model incorporating risk aversion and aggregate uncertainty. We show that introducing risk aversion in a model with several states of the world can also yield several rating classes. In this case our previous analysis can be interpreted as determining the optimal cutoff of the lowest investment grade rating class (e.g. BBB). Second, we provide a numerical example to show that the effects of the moments of the expected quality distribution on the optimal cutoff have the same sign as before even with risk aversion and several rating classes.

We provide the simplest possible setup which is rich enough to illustrate the idea. Assume that buyers are risk-averse. Their utility of an asset is equal to t but their expected utility depends on both the mean and the variance of the quality of the asset they buy. We include a second mass point at $\bar{t}_2$, $\bar{t}_2 \geq \bar{t}$, with mass γ_i in state i. To avoid confusion denote $\bar{t}$ by $\bar{t}_1$. See Fig. 4.6.

If buyers are risk-averse, a welfare maximizing rating strategy needs to perfectly disclose the type of all assets with a positive value because any kind of pooling and being vague about a firm's quality leads to a welfare loss. However, such a strategy cannot be optimal for the

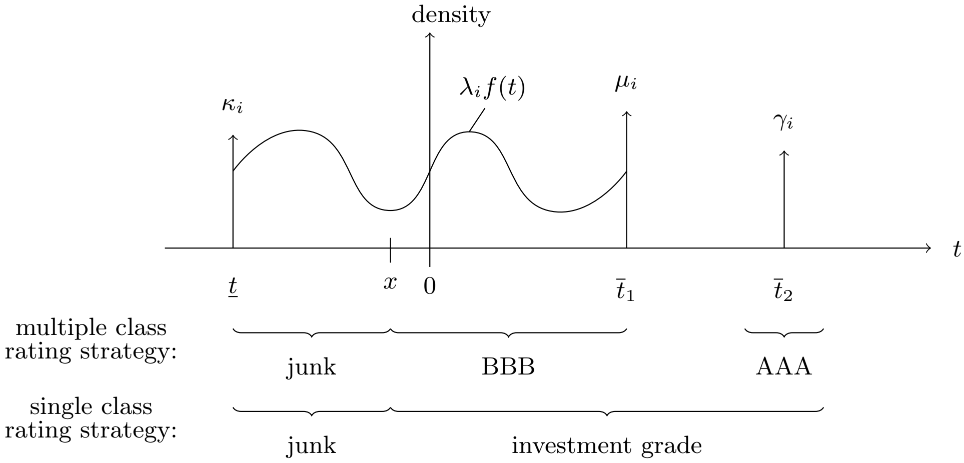

Figure 4.6: κ_i, μ_i and γ_i are the mass points at $\underline{t}$, $\bar{t}_1$ and $\bar{t}_2$ in state i. λ_i is the mass in state i that is allotted to the types $t \in (\underline{t}, \bar{t}_1)$ with the distribution F.

rating agency.[17] To analyze a general model with risk-aversion is beyond the scope of this article. In the following, we compare two rating strategies: (i) pooling all types above a cutoff in one rating class, which is the optimal strategy without risk-aversion, and (ii) a strategy in which the agency only pools low types and rates high types separately. Doherty et al. (2012) show that strategy (ii) is optimal in a model with one state of the world if the level of risk-aversion is sufficiently high.

Analogously to the case without risk-aversion, we derive the profit of the agency if it pools all types above a cutoff x in one class. The expected type above a cutoff x is

$$Q_i(x) := E[t|t \geq x]$$

and the variance is

$$\sigma_i(x) := \mathrm{Var}[t|t \geq x].$$

The buyer's valuation for the asset of a seller in this rating class is

$$Q_i(x) - a\sigma_i(x)$$

where a is a measure for risk-aversion. If $a = 0$, the buyers are risk-neutral and the model is equivalent to before.

[17]To ensure that all firms with $t \geq 0$ are willing to pay the rating fee under full disclosure, the rating fee has to be 0.

The profit of the agency if it pools all types is

$$\begin{aligned}\hat{\Pi}(x) :=& \left(\sum_i (\lambda_i(1-F(x)) + \mu_i + \gamma_i)\epsilon_i\right)\left(\sum_i \hat{\epsilon}_i(Q_i(x) - a\sigma_i(x))\right)\\ =&(1-F(x)+\tilde{\mu}+\tilde{\gamma})\sum_i \hat{\epsilon}_i(Q_i(x) - a\sigma_i(x))\end{aligned}$$

where $\tilde{\gamma}$ is the expected value of γ, $\tilde{\gamma} = \sum_i \epsilon_i \gamma_i$.

Alternatively, the rating agency can pool $t \in [x, \bar{t}_1]$ and rate $\bar{t}_2$ separately as shown in Fig. 4.6. If the agency rates types $\bar{t}_2$ in a separate class, these sellers are willing to pay a high rating fee (up to $\bar{t}_2$) and therefore the rating fee is determined by sellers in the class $t \in [x, \bar{t}_1]$. Keeping the cutoff x constant, the mass of rated firms is the same for both strategies and the rating fee decides which rating strategy yields higher profits. If the agency pools types $t \in [x, \bar{t}_1]$, the expected type in this rating class is smaller than $Q_i(x)$ but the variance is also smaller than $\sigma_i(x)$. Thus, it is not straight forward to see under which strategy the rating fee can be higher.

Now, we derive sufficient conditions such that the agency prefers to rate $\bar{t}_2$ separately instead of pooling all types above x in one rating class. Define $z_i := \gamma_i \bar{t}_2$ and $\tilde{z}$ as the expected value of z_i, $\tilde{z} := \sum_i \epsilon_i z_i$. Rewrite $\tilde{z}$ as $\tilde{z} = \bar{t}_2 \tilde{\gamma}$, which can be interpreted as the agency's profit if it charges a rating fee of $\bar{t}_2$ and rates only firms with type $\bar{t}_2$. Remember that $\Pi(x)$ is defined as the profit if the agency rates only $t \in [x, \bar{t}_1]$ and pools them all in one class.

Proposition 4.6. *Take an arbitrary cutoff x. For any $\tilde{z}$ with $\tilde{z} \leq \Pi(x)$ there exists a $\overline{T}_2$ such that for all $\bar{t}_2 \geq \overline{T}_2$ the rating agency is better off pooling $t \in [x, \bar{t}_1]$ and rating $\bar{t}_2$ in a separate class than pooling all types above x in one rating class.*

Since investors are risk-averse, their expected utility buying an asset in a given rating class decreases if the variance inside this rating class becomes larger. If the variance is sufficiently large, investors are not willing to pay any positive price for an asset even if the expected quality is positive. Thus, if the variance is large, the agency is better off splitting the types in several rating classes in order to reduce the variance inside one class and to increase investors' willingness to pay for an asset. The condition that $\tilde{z} \leq \Pi(x)$ ensures that the agency does not prefer to charge a rating fee of $\bar{t}_2$ and to exclude firms with $t < \bar{t}_2$ from the rating.

Risk aversion does not only have the effect of multiple rating classes becoming optimal, but it also has an additional effect on the optimal cutoff. Increasing the cutoff reduces the variance in a rating class and this can give additional incentives to increase the cutoff.[18] In the following we provide numerical examples in which we show that the effects of the first, second, and higher moments are similar to our analysis without risk aversion.[19] In the numerical example we have four states of the world. We take the Generalized Pareto distribution $F(t) = 1 - ((1-t)/2)^3$ for $t \in (-1, 1)$ and fix $\bar{t}_1 = 1$. This gives us $E_0 = 1/4$. We fix $\lambda = 5$, $\bar{t}_2 = 110$, and $\nu_i = 0.0001$ for all i. The states only differ in the weights μ_i at the mass point at $\bar{t}_1$, with $\mu_1 = 0.03$, $\mu_2 = 0.2$, $\mu_3 = 0.4$, and $\mu_4 = 0.7$. Changing the moments of the aggregate distribution, we keep the distribution inside a state constant (and therefore also the expected type) and only vary the probabilities for the different states. There is a one-to-one mapping from $(\epsilon_1, \epsilon_2, \epsilon_3)$ to (m_1, m_2, m_{3+}) and the fourth probability is pinned down by $\epsilon_4 = 1 - \epsilon_1 - \epsilon_2 - \epsilon_3$. For all values of the example, the probabilities are in $[0, 1]$.

Figures 4.7, 4.8, and 4.9 illustrate the change of the optimal cutoff as m_{3+}, m_1 and m_2 are changed while keeping the other moments constant. The solid line is the optimal cutoff for $a = 0$, the dashed line for $a = 0.01$ and the dotted-dashed line for $a = 0.02$. If investors are risk neutral, $a = 0$, the agency pools all types above the cutoff in one class. For $a = 0.01$ and $a = 0.02$, investors are risk averse and the agency prefers to pool all types $t \in [x, \bar{t}_1]$ in one class and to rate $\bar{t}_2$ separately. Note that increasing the level of risk aversion leads to an increase in the optimal cutoff x^*. The figures show that our results from the main part of the paper carry over to a setup including risk-aversion: Keeping the other moments constant, a higher mean, a lower variance, or an increase in the higher order skewness lead to an increase in the optimal cutoff. For changes with the opposite sign, the optimal cutoff decreases.

4.7 Conclusions

We have considered the profit maximizing rating strategy of a rating agency in the face of aggregate uncertainty. We have shown that with risk neutral investors it is still optimal for the rating agency, as in a setup without aggregate uncertainty, to choose only one rating class for rated firms and to not rate the remaining firms.

The model's predictions about the cutoff for the rating class strikingly differ from the

[18]Doherty et al. (2012) show that the optimal cutoff can be positive even with only one state of the world if the level of risk aversion is sufficiently high.

[19]In the main part of the paper the moments were defined for the distribution of the expected type in $[0, \bar{t}]$ (scaled by $\bar{t}$). For the sake of comparison, in the numerical examples the moments are defined for the distribution of the expected type in $[x, \bar{t}_1]$ and thus, the expected type is not influenced by changes in the mass on $\bar{t}_2$. We deviate from our previous analysis by taking the threshold x as the lower bound of the interval. In this way we can determine the optimal cutoff explicitly and not only its sign.

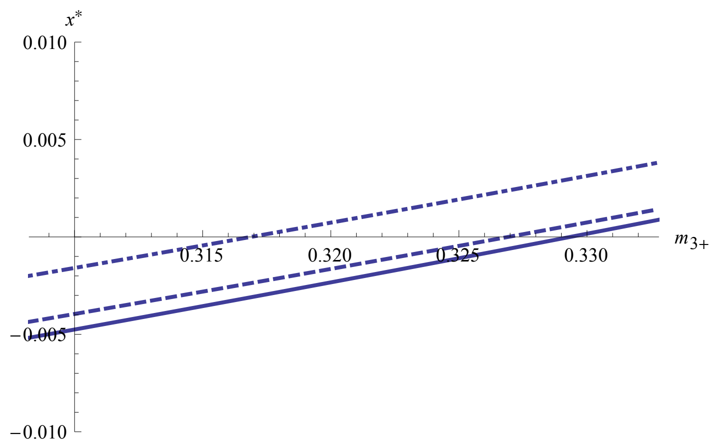

Figure 4.7: Values of the optimal threshold x^* as m_{3+} is changed and m_1 and m_2 are kept constant. For the solid line $a = 0$, for the dashed line $a = 0.01$ and for the dotted-dashed line $a = 0.02$. The rating strategy for the solid line is to pool all types above x. For the dashed and dotted-dashed line all types in $[x, \bar{t}_1]$ are pooled and $\bar{t}_2$ is rated separately. (The starting values are $\epsilon_i = 1/4$ for all i. This implies $m_1 = 0.47627$, $m_2 = 0.244859$ and as a starting value $m_{3+} = 0.321538$.)

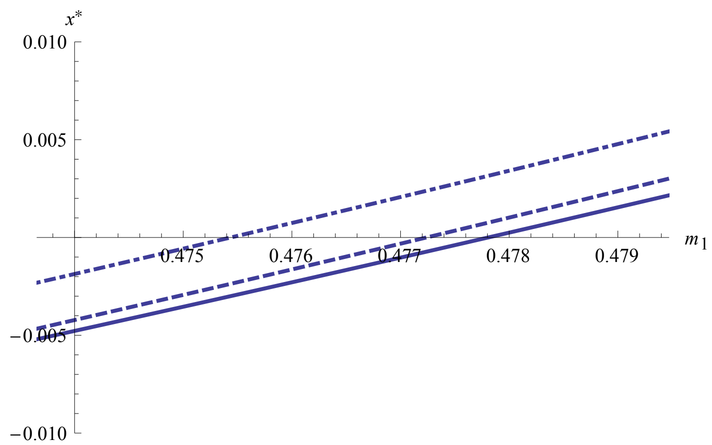

Figure 4.8: Values of the optimal threshold x^* as m_1 is changed and m_2 and m_{3+} are kept constant. For the solid line $a = 0$, for the dashed line $a = 0.01$ and for the dotted-dashed line $a = 0.02$. The rating strategy for the solid line is to pool all types above x. For the dashed and dotted-dashed line all types in $[x, \bar{t}_1]$ are pooled and $\bar{t}_2$ is rated separately. (The starting values are $\epsilon_i = 1/4$ for all i. This implies $m_2 = 0.244859$, $m_{3+} = 0.321538$ and as a starting value $m_1 = 0.47627$.)

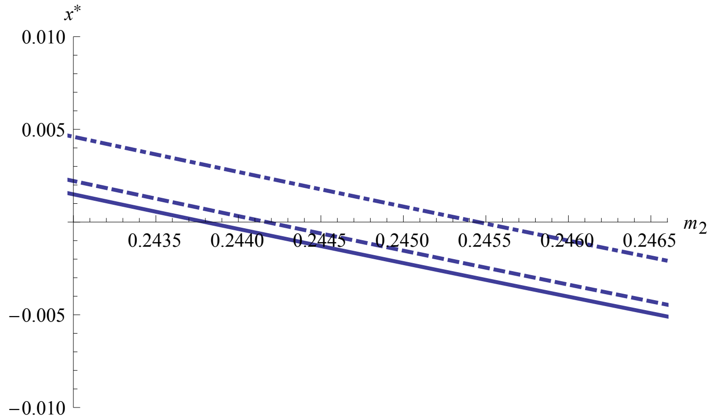

Figure 4.9: Values of the optimal threshold x^* as m_2 is changed and m_1 and m_{3+} are kept constant. For the solid line $a = 0$, for the dashed line $a = 0.01$ and for the dotted-dashed line $a = 0.02$. The rating strategy for the solid line is to pool all types above x. For the dashed and dotted-dashed line all types in $[x, \bar{t}_1]$ are pooled and $\bar{t}_2$ is rated separately. (The starting values are $\epsilon_i = 1/4$ for all i. This implies $m_1 = 0.47627$, $m_{3+} = 0.321538$ and as a starting value $m_2 = 0.244859$.)

predictions of a model without aggregate uncertainty: the rating agency has more of an incentive to be too lenient if the expected average quality is small, the variance large, and the higher order skewness small. For larger averages, smaller variances, and larger higher order skewness the opposite holds: the rating agency has more of an incentive to be too strict. These results can be interpreted as ratings having either a pro-cyclical or an anti-cyclical effect. We outline an empirical strategy to estimate the moments of aggregate uncertainty which can be used to determine which effect dominates.

Our analysis identifies one up to now unconsidered factor that affects the rating strategy of an agency – aggregate uncertainty – and thereby sheds further light in understanding the behavior of rating agencies. In line with our model, one disturbing effect of using ratings as the basis for financial regulation is that a possible pro-cyclicality of ratings leads to a pro-cyclicality of capital adequacy requirements for banks, and hence to a pro-cyclicality of lending. One solution is to avoid using ratings as the basis for financial regulation. Another is to counterbalance the pro-cyclicality of ratings by adding counter-cyclicality to capital adequacy requirements that are based on ratings.

The usual disclaimer for the policy implications holds. This article is about a thorough analysis of the effects of aggregate uncertainty, shutting down other effects such as like reputation cycles, imperfect rating technology, and competition between agencies. Further, the implications of the theory depend on the empirical moments of the distribution of aggregate uncertainty. Hence, an empirical analysis is needed to estimate these moments and the relative magnitude of the different effects. Our article provides a starting point for such an empirical analysis. This paper also serves as a word of caution: using a distribution which is pinned down by its mean and variance (e.g. a normal distribution) for an empirical analysis will neglect the impact of the higher order skewness. However, the skewness is crucial for the incentive of the rating agency to distort ratings.

4.8 Appendix

Proof of Lemma 4.1

Proof. (i) Denote the class containing $\underline{t}$ as T_n. Define an alternative rating class $T_n^* = \{\underline{t}\} \cup [0, \bar{t}]$. Observe $\kappa_i \underline{t} + \lambda_i \int_0^{\bar{t}} t dF(t) + \mu_i \bar{t}$ is the expected quality in T_n^* in state i, $E\left[T_{n,i}^*\right]$, and by Assumption 1 this is smaller than zero. The expected quality in class T_n in state i, $E[T_{n,i}]$, can be larger or smaller than $E\left[T_{n,i}^*\right]$. If $E[T_{n,i}]$ is smaller than $E\left[T_{n,i}^*\right]$, it follows directly that $E[T_{n,i}] < 0$.

If $E\left[T_{n,i}\right]$ is larger than $E\left[T_{n,i}^*\right]$, T_n must include types $t \in [E\left[T_{n,i}^*\right], 0]$ to raise the expected quality. Including negative qualities in T_n can increase the expected type in comparison to T_n^* but the expected type $E[T_{n,i}]$ stays negative. Therefore, the willingness to pay for a rating in category T_n is negative and the rating agency prefers not to have category T_n.

(ii) Take a rating strategy $\{\tilde{T}_m\}_{m=1}^{\tilde{M}}$. Assume that $\bar{t}$ is not in any $\tilde{T}_m$. Define for all rating classes $\tilde{T}_m$ the expected value

$$E_m^* = \frac{\int_{t \in \tilde{T}_m} t dF(t)}{\int_{t \in \tilde{T}_m} dF(t)} \tag{4.10}$$

which is constant over all states of the world. The price is determined by the lowest willingness to pay $\min_m E_m^*$ and the expected mass of rated firms $\sum_i \epsilon_i \sum_m \int_{t \in \tilde{T}_m} \lambda_i dF(t)$. Using $\sum_i \epsilon_i \lambda_i = 1$, we get for profits

$$\tilde{\Pi} = \left[\min_m E_m^*\right] \left\{ \sum_{m=1}^{\tilde{M}} \int_{t \in \tilde{T}_m} dF(t) \right\} \tag{4.11}$$

Now take a rating strategy with $M = \tilde{M}+1$, $T_m = \tilde{T}_m$ for $m \leq \tilde{M}$ and $T_M = \{\bar{t}\}$. Including the $\bar{t}$ types adds expected mass $\tilde{\mu}$ to the mass of rated firms. Hence, expected profits are

$$\Pi = \left[\min\left(\{\bar{t}\} \cup \{E_m^*\}_{m=1}^{\tilde{M}}\right)\right] \left\{ \tilde{\mu} + \sum_{m=1}^{\tilde{M}} \int_{t \in T_m} dF(t) \right\} \tag{4.12}$$

Since in (4.12) the expression in square brackets is weakly greater than in (4.11) and the expression in curly braces is strictly greater in (4.12) than in (4.11), we have $\Pi > \tilde{\Pi}$. $\square$

Proof of Lemma 4.2

Proof. It holds that

$$\sum_i \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) = 0$$

because of $\sum_i \epsilon_i \mu_i / \tilde{\mu} = \sum_i \epsilon_i \lambda_i / \tilde{\lambda} = 1$ Define two sets of states of the world; $i \in A$ if $\frac{\mu_i \epsilon_i}{\lambda_i} > \tilde{\mu}$ and $i \in B$ if $\frac{\mu_i \epsilon_i}{\lambda_i} \leq \tilde{\mu}$. Thus

$$\sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) + \sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) = 0$$

and multiply by a constant c

$$\sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c + \sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c = 0. \tag{4.13}$$

The expected quality in state i is

$$E_i = \frac{\lambda_i \int_{t \in T} t dF + \mu_i \bar{t}}{\lambda_i \int_{t \in T} dF + \mu_i} = \frac{\int_{t \in T} t dF + \mu_i / \lambda_i \bar{t}}{\int_{t \in T} dF + \mu_i / \lambda_i}$$

and is increasing in μ_i / λ_i. Define c as $\frac{\int_{t \in T} t dF + \tilde{\mu} \bar{t}}{\int_{t \in T} dF + \tilde{\mu}}$. The expected quality E_i for $i \in A$ is larger than c and $E_i < c$ for $i \in B$. It follows that

$$\sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c < \sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i \tag{4.14}$$

and

$$\sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c < \sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i. \tag{4.15}$$

Using inequalities (4.14) and (4.15) gives us

$$\begin{aligned} &\sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i + \sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) E_i \\ > &\sum_{i \in A} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c + \sum_{i \in B} \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} - \frac{\epsilon_i \lambda_i}{\tilde{\lambda}} \right) c \end{aligned}$$

which is equal 0 by equation (4.13). Therefore, it holds that

$$\sum_i \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} E_i\right) - \sum_i \left(\frac{\epsilon_i \lambda_i}{\tilde{\lambda}} E_i\right) > 0.$$

Since $\sum_i \left(\frac{\epsilon_i \mu_i}{\tilde{\mu}} E_i\right)$ is the willingness to pay for a rating for type $\bar{t}$ and $\sum_i \left(\frac{\epsilon_i \lambda_i}{\tilde{\lambda}} E_i\right)$ for type $t \in (\underline{t}, \bar{t})$, the lemma follows. □

Proof of Lemma 4.3

Proof. Label the rating class that includes $\bar{t}$ as $\tilde{T}_1$ and the remaining rating classes as $\tilde{T}_{-1} = \cup_{m \neq 1} \tilde{T}_m$. Denote the expected type of $\tilde{T}_1$ conditional on being in state i as $\tilde{E}_i = [\int_{t \in \tilde{T}_1} t dF(t) + \mu_i \bar{t}]/[\int_{t \in \tilde{T}_1} dF(t) + \mu_i]$. Denote the mass of all other classes as $\mu^* = \int_{t \in \tilde{T}_{-1}} dF(t)$ and the expected type as $t^* = [\int_{t \in \tilde{T}_{-1}} t dF(t)]/[\int_{t \in \tilde{T}_{-1}} dF(t)]$.

Profits for only one rating class $T_1 = \cup_m \tilde{T}_m$ are

$$\Pi = \left[\sum_i \hat{\epsilon}_i E_i\right] (\mu^* + \tilde{\mu})$$

where

$$E_i = \frac{\lambda_i (\int_{t \in \tilde{T}_{-1}} t dF(t)) + \int_{t \in \tilde{T}_1} t dF(t) + \mu_i \bar{t}}{\lambda_i (\int_{t \in \tilde{T}_{-1}} dF(t) + \int_{t \in \tilde{T}_1} dF(t)) + \mu_i}$$

is the expected type in state i if there is only one rating class. Profits for separate rating classes $\{\tilde{T}_m\}$ are

$$\tilde{\Pi} = \left[\min\left(\{E_m^*\}_{m=1}^{\tilde{M}} \cup \left\{\sum_i \hat{\epsilon}_i \tilde{E}_i\right\}\right)\right] (\mu^* + \tilde{\mu}),$$

where E_m^* is defined as in (4.10). Further, define the profit in case all rating classes $m \neq 1$ were merged, such that one had two rating classes $\tilde{T}_1$ and $\cup_{m=2}^{\tilde{M}} \tilde{T}_m$, as

$$\hat{\Pi} = \left[\min\left\{t^*, \sum_i \hat{\epsilon}_i \tilde{E}_i\right\}\right] (\mu^* + \tilde{\mu}).$$

Since t^* is a weighted average of $\{\tilde{E}_m\}_{m=1}^{\tilde{M}}$, we have $t^* \geq \min\{\tilde{E}_m\}_{m=1}^{\tilde{M}}$ and therefore $\hat{\Pi} \geq \tilde{\Pi}$. (Note that Π, $\hat{\Pi}$, and $\tilde{\Pi}$ only differ in the expressions in square brackets.)

We will prove the lemma by contradiction. Assume to the contrary that separate classes are desirable, i.e. $\tilde{\Pi} > \Pi$. This implies $\hat{\Pi} > \Pi$, which is equivalent to

$$\min\left\{t^*, \sum_i \hat{\epsilon}_i \tilde{E}_i\right\} > \sum_i \hat{\epsilon}_i E_i,$$

by comparison of the expressions in square brackets. This condition is equivalent to both

$$t^* > \sum_i \hat{\epsilon}_i E_i \tag{4.16}$$

and

$$\sum_i \hat{\epsilon}_i \tilde{E}_i > \sum_i \hat{\epsilon}_i E_i \tag{4.17}$$

being satisfied at the same time.

The expected value E_i can be written as weighted average of t^* and $\tilde{E}_i$ for every state i

$$\begin{aligned} E_i =& \frac{\lambda_i(\int_{t\in\tilde{T}_{-1}} tdF(t) + \int_{t\in\tilde{T}_1} tdF(t)) + \mu_i\bar{t}}{\lambda_i(\int_{t\in\tilde{T}_{-1}} dF(t) + \int_{t\in\tilde{T}_1} dF(t)) + \mu_i} \\ =& \frac{\lambda_i t^* \int_{t\in\tilde{T}_{-1}} dF(t) + \tilde{E}_i(\int_{t\in\tilde{T}_1} \lambda_i dF(t) + \mu_i)}{\lambda_i(\int_{t\in\tilde{T}_{-1}} dF(t) + \int_{t\in\tilde{T}_1} dF(t)) + \mu_i}. \end{aligned}$$

Solving for $\tilde{E}_i$, we get

$$\tilde{E}_i = E_i + \frac{\lambda_i \int_{t\in\tilde{T}_{-1}} dF(t)}{\lambda_i \int_{t\in\tilde{T}_1} dF(t) + \mu_i}(E_i - t^*)$$

Plugging $\tilde{E}_i$ into (4.17), we get

$$\sum_i \hat{\epsilon}_i \left(E_i + \frac{\lambda_i \int_{t\in\tilde{T}_{-1}} dF(t)}{\lambda_i \int_{t\in\tilde{T}_1} dF(t) + \mu_i}(E_i - t^*)\right) > \sum_i \hat{\epsilon}_i E_i$$

or equivalently

$$\sum_i \hat{\epsilon}_i \left(\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + \mu_i/\lambda_i}(E_i - t^*)\right) > 0. \tag{4.18}$$

Define two sets of states of the world; $i \in A$ if $E_i \geq t^*$ and $i \in B$ if $E_i < t^*$. It holds that $\mu_i/\lambda_i > \mu_j/\lambda_j$ for all $i \in A$ and $j \in B$. This can be seen by checking that

$$E_i = \frac{(\int_{t\in\tilde{T}_{-1}} tdF(t) + \int_{t\in\tilde{T}_1} tdF(t)) + \mu_i/\lambda_i\bar{t}}{(\int_{t\in\tilde{T}_{-1}} dF(t) + \int_{t\in\tilde{T}_1} dF(t)) + \mu_i/\lambda_i}$$

is increasing in μ_i/λ_i. Denote $c_A = \min\{\mu_i/\lambda_i | i \in A\}$ and $c_B = \max\{\mu_i/\lambda_i | i \in B\}$. Note that $c_A > c_B$.

Then (4.16) can be rewritten as

$$\sum_i \hat{\epsilon}_i(E_i - t^*) < 0$$

which is equivalent to

$$\sum_{i\in A} \hat{\epsilon}_i(E_i - t^*) + \sum_{i\in B} \hat{\epsilon}_i(E_i - t^*) < 0.$$

This implies

$$\left[\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_A} \sum_{i\in A} \hat{\epsilon}_i(E_i - t^*)\right] + \left[\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_B} \sum_{i\in B} \hat{\epsilon}_i(E_i - t^*)\right] < 0 \qquad (4.19)$$

since

$$\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_A} < \frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_B}$$

and the sum over $i \in A$ is positive and the sum over $i \in B$ is negative. Since $\mu_i/\lambda_i \geq c_A$ for all $i \in A$ and $\sum_{i\in A} \hat{\epsilon}_i(E_i - t^*)$ positive, the first expression in square brackets in (4.19) is bounded form below by

$$\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_A} \sum_{i\in A} \hat{\epsilon}_i(E_i - t^*) \geq \sum_{i\in A} \hat{\epsilon}_i \left(\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + \mu_i/\lambda_i}(E_i - t^*)\right). \qquad (4.20)$$

The second expression in square brackets is bounded from below by

$$\sum_{i\in B} \hat{\epsilon}_i(E_i - t^*) \frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + c_B} \geq \sum_{i\in B} \hat{\epsilon}_i \frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + \mu_i/\lambda_i}(E_i - t^*). \qquad (4.21)$$

because of $\mu_i/\lambda_i \leq c_B$ for all $i \in B$ and the negativity of $\sum_{i\in B} \hat{\epsilon}_i(E_i - t^*)$.

(4.19),(4.20) and (4.21) imply

$$\sum_{i\in A} \hat{\epsilon}_i \left(\frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + \mu_i/\lambda_i}(E_i - t^*)\right) + \sum_{i\in B} \hat{\epsilon}_i \frac{\int_{t\in\tilde{T}_{-1}} dF(t)}{\int_{t\in\tilde{T}_1} dF(t) + \mu_i/\lambda_i}(E_i - t^*) < 0$$

which contradicts (4.18). $\square$

Proof of Lemma 4.4

Proof. Assume that $\hat{T}$ is not convex. Take a convex set T' such that it has the same expected mass of rated firms ($\int_{t\in\hat{T}} dF(t) = \int_{t\in T'} dF(t)$) and both sets include $\bar{t}$. Remember that the profit is

$$\Pi(T) = \left[\sum_i \hat{\epsilon}_i \frac{\lambda_i \int_{t\in T} t dF(t) + \mu_i \bar{t}}{\lambda_i \int_{t\in T} dF(t) + \mu_i}\right] \left(\int_{t\in T} dF(t) + \tilde{\mu}\right).$$

Since $\int_{t\in\hat{T}} dF(t) = \int_{t\in T'} dF(t)$, comparing the profits $\Pi(\hat{T})$ and $\Pi(T')$ boils down to comparing the willingness to pay for $\hat{T}$ and T', which is given in square brackets. Since $\hat{T}$ is not convex, there is at least one unrated hole in the middle and it is possible to rate the mass in the holes instead of rating some types below with the same mass. This increases $\int_{t\in T} t dF(t)$, while the mass of rated types stays constant. It follows that $\frac{\lambda_i \int_{t\in T} t dF(t)+\mu_i \bar{t}}{\lambda_i \int_{t\in T} dF(t)+\mu_i}$ is greater for T' than for $\hat{T}$ and hence $\Pi(T') > \Pi(\hat{T})$. Therefore, it is optimal to rate a set T which is convex and includes $\bar{t}$. □

Proof of Lemma 4.5

Proof. Analogously to the proof of Lemma 4.2 define two sets of states of the world; $i \in A$ if $\frac{\mu_i}{\lambda_i} > \frac{\tilde{\mu}}{\tilde{\lambda}}$ and $i \in B$ if $\frac{\mu_i}{\lambda_i} \leq \frac{\tilde{\mu}}{\tilde{\lambda}}$. It holds that $\sum_i \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) = 0$ which we can write as $\sum_A \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) + \sum_B \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) = 0$. Multiplied by a constant $c = \frac{1}{(1-F)+\tilde{\mu}}$ the expression is still equal to 0. For $i \in A$, $\frac{1}{(1-F)+\mu_i/\lambda_i}$ is smaller than $\frac{1}{(1-F)+\tilde{\mu}}$ and for $i \in B$ it is the other way round. It follows that

$$\sum_A \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) \frac{1}{(1-F) + \mu_i/\lambda_i} + \sum_B \epsilon_i (\mu_i - \lambda_i \tilde{\mu}) \frac{1}{(1-F) + \mu_i/\lambda_i} < 0$$

because $\mu_i - \lambda_i \tilde{\mu}$ is positive for $i \in A$ and negative for $i \in B$. This is equivalent to $\frac{(\bar{t}-E_0)(1-F)}{\tilde{\lambda}(1-F)+\tilde{\mu}} \sum_i \hat{\epsilon}_i \frac{\tilde{\lambda}\mu_i - \tilde{\mu}\lambda_i}{\lambda_i(1-F)+\mu_i} < 0$ and thus, $\sum_i \hat{\epsilon}_i \frac{\lambda_i(1-F)E_0+\mu_i \bar{t}}{\lambda_i(1-F)+\mu_i} < \frac{\tilde{\lambda}(1-F)E_0+\tilde{\mu}\bar{t}}{\tilde{\lambda}(1-F)+\tilde{\mu}}$. □

Proof of Proposition 4.2

Proof.

$$\begin{aligned}
L'(x) =& \Pi'(x) - W'(x) \\
=& f(x)\left((\hat{E} - \frac{1-F(x)+\hat{\mu}}{f}\hat{E}') - (\tilde{E} - \frac{1-F(x)+\tilde{\mu}}{f}\tilde{E}')\right) \\
=& f(x)\left(\hat{E} - \tilde{E} + \frac{1-F(x)+\tilde{\mu}}{f}(\tilde{E}' - \hat{E}')\right).
\end{aligned} \tag{4.22}$$

We know that $\hat{E} \geq \tilde{E}$ but the sign of $\tilde{E}' - \hat{E}'$ can go in both directions.

Next, we rewrite (4.22) such that we can express $L'(x)$ only in terms of E_i, $\tilde{E}$, and $\hat{E}$. The derivative of $\tilde{E}$ with respect to x is

$$\sum_i \hat{\epsilon}_i \frac{\partial E_i}{\partial x} = \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1-F(x)) + \mu_i} f(x)\lambda_i.$$

and analogously it can be shown that

$$\frac{\partial \hat{E}}{\partial x} = \frac{\hat{E} - x}{1 - F(x) + \tilde{\mu}} f(x).$$

Using these two expressions in (4.22), we can write

$$\begin{aligned}
L'(x) =& f(x)\left(\hat{E} - \tilde{E} + \frac{1-F(x)+\tilde{\mu}}{f}\left(\sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1-F(x)) + \mu_i}\lambda_i f - \frac{\hat{E} - x}{1-F(x)+\tilde{\mu}} f\right)\right) \\
=& f(x)\left(\hat{E} - \tilde{E} + (1 - F(x) + \tilde{\mu}) \sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1-F(x)) + \mu_i}\lambda_i - (\hat{E} - x)\right).
\end{aligned}$$

From the definitions of E_i and $\hat{E}$ we derive $\mu_i = \lambda_i(1-F(x))\frac{E_i - E_0}{\bar{t} - E_i}$ and $\tilde{\mu} = (1-F(x))\frac{\hat{E} - E_0}{\bar{t} - \hat{E}}$ which leads to

$$\begin{aligned}
& L'(x) \\
=& f(x)\left(x - \tilde{E} + \left(1 - F(x) + (1-F(x))\frac{\hat{E} - E_0}{\bar{t} - \hat{E}}\right)\sum_i \hat{\epsilon}_i \frac{E_i - x}{\lambda_i(1-F(x)) + \lambda_i(1-F(x))\frac{E_i - E_0}{\bar{t} - E_i}}\right) \\
=& f(x)\left(x - \tilde{E} + (1 + \frac{\hat{E} - E_0}{\bar{t} - \hat{E}})\sum_i \hat{\epsilon}_i \frac{(E_i - x)(\bar{t} - E_i)}{\bar{t} - E_0}\right) \\
=& f(x)\left(x - \tilde{E} + \sum_i \hat{\epsilon}_i \frac{(E_i - x)(\bar{t} - E_i)}{\bar{t} - \hat{E}}\right).
\end{aligned}$$

Remember that $W'(0) = 0$ which implies that $L'(0) = \Pi'(0)$. Thus, the sign of $L'(0)$ determines the sign of the profit maximizing cutoff. To determine the sign of $L'(0)$ we set $x = 0$ in the above expression and

$$\begin{aligned}
L'(0) =& f(0)\left(-\tilde{E} + \frac{\sum_i \hat{\epsilon}_i E_i(\bar{t} - E_i)}{\bar{t} - \hat{E}}\right) \\
=& f(0)\left(-\tilde{E} + \frac{\bar{t}\tilde{E} - \sum_i \hat{\epsilon}_i E_i^2 - \tilde{E}\hat{E} + \tilde{E}\hat{E}}{\bar{t} - \hat{E}}\right) \\
=& f(0)\left(-\tilde{E} + \frac{\tilde{E}\hat{E} - \sum_i \hat{\epsilon}_i E_i^2}{\bar{t} - \hat{E}} + \frac{(\bar{t} - \hat{E})\tilde{E}}{\bar{t} - \hat{E}}\right) \\
=& f(0)\left(\underbrace{-\tilde{E}}_{marginal\ effect} + \underbrace{\frac{\tilde{E}}{\bar{t} - \hat{E}}\left(\hat{E} - \frac{\sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}\right) + \tilde{E}}_{inframarginal\ effect}\right).
\end{aligned}$$

This expression gives us another way to write the inframarginal effect of a change of the threshold at $x = 0$ on the profit Π. We can simplify $L'(0)$ to

$$\begin{aligned}
L'(0) =& \frac{f(0)\tilde{E}}{\bar{t} - \hat{E}}\left(\hat{E} - \frac{\sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}\right) \\
=& \frac{f(0)\tilde{E}}{\bar{t} - \hat{E}}\left(\hat{E} - \tilde{E} - \frac{(\sum_i \hat{\epsilon}_i E_i)^2 - \sum_i \hat{\epsilon}_i E_i^2}{\tilde{E}}\right).
\end{aligned}$$

□

Proof of Proposition 4.5

Proof. The expression for ϵ_N can be derived by observing that

$$O_{N-1} = \sum_{i=N}^{N} \epsilon_i(E_i - y_{N-1}) = \epsilon_n(y_N - y_{N-1}).$$

The expression for ϵ_j for $1 < j < N$ can be obtained by first observing that

$$O_{j-1} - O_j = \sum_{i=j}^{N} \epsilon_i(E_i - E_{j-1}) - \sum_{i=j+1}^{N} \epsilon_i(E_i - E_j) = \sum_{i=j}^{N} \epsilon_i(E_j - E_{j-1}),$$

where the first equality follows from (4.8) and the second equality can be obtained by rearranging the sums. Dividing by $E_j - E_{j-1}$ yields

$$\frac{O_{j-1} - O_j}{E_j - E_{j-1}} = \sum_{i=j}^{N} \epsilon_i,$$

and taking differences

$$\frac{O_j - O_{j+1}}{y_{j+1} - y_j} - \frac{O_{j-1} - O_j}{y_j - y_{j-1}} = \sum_{i=j+1}^{N} \epsilon_i - \sum_{i=j}^{N} \epsilon_i = \epsilon_j,$$

that is, the expression for ϵ_j for $1 < j < N$ in the proposition. The expression for ϵ_1 simply follows from that fact that probabilities add up to one. $\square$

Proof of Proposition 4.6

Proof. We want to analyze the effect of a change of $\bar{t}_2$, while keeping $z_i = \gamma_i \bar{t}_2$ constant. For this purpose the variance $\sigma_i(x)$ can be rewritten as

$$\begin{aligned}
\sigma_i(x) =& \frac{\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + \gamma_i \bar{t}_2^2}{\lambda_i(1 - F(x)) + \mu_i + \gamma_i} - \left(\frac{\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + \gamma_i \bar{t}_2}{\lambda_i(1 - F(x)) + \mu_i + \gamma_i} \right)^2 \\
=& \frac{(\lambda_i(1 - F(x)) + \mu_i + \gamma_i)(\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + \gamma_i \bar{t}_2^2) - (\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + \gamma_i \bar{t}_2)^2}{(\lambda_i(1 - F(x)) + \mu_i + \gamma_i)^2} \\
=& \frac{(\lambda_i(1 - F(x)) + \mu_i + z_i/\bar{t}_2)(\lambda_i \int_x^{\bar{t}_1} t^2 dF(t) + \mu_i \bar{t}_1^2 + z_i \bar{t}_2) - (\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + z_i)^2}{(\lambda_i(1 - F(x)) + \mu_i + z_i/\bar{t}_2)^2}
\end{aligned}$$

For $\bar{t}_2 \to \infty$ we get that the variance $\sigma_i(x)$ goes to infinity

$$\lim_{\bar{t}_2 \to \infty} \sigma_i(x) = \infty$$

and the expected type $Q_i(x)$ converges to

$$\lim_{\bar{t}_2 \to \infty} Q_i(x) = \frac{\lambda_i \int_x^{\bar{t}_1} t dF(t) + \mu_i \bar{t}_1 + z_i}{\lambda_i(1 - F(x)) + \mu_i} < \infty.$$

It follows that for $a > 0$, the utility in state i, $Q_i(x) - a\sigma_i(x)$, becomes negative if $\bar{t}_2$ is large enough. This implies that buyers are not willing to pay a positive price for a rated firm if the variance of types in one rating class is too high. Thus, for every cutoff x there is a $\bar{t}_2$ large enough such that the rating agency is better off pooling $t \in [x, \bar{t}_1]$ and rating $\bar{t}_2$ in a separate class than pooling all types above x.

A further condition that needs to be satisfied is that the agency does not prefer to charge a rating fee of $\bar{t}_2$ and to rate only firms with type $\bar{t}_2$ which yields profits of $\bar{t}_2\tilde{\gamma}$. Since we keep $\gamma_i\bar{t}_2$ constant when we increase $\bar{t}_2$, the profit of only rating types $\bar{t}_2$ stays constant. A sufficient condition such that the agency prefers to rate $[x, \bar{t}_1]$ and $\bar{t}_2$ is that the profit of pooling $t \in [x, \bar{t}_1]$ and not rating $\bar{t}_2$ is larger than the profit of only rating $\bar{t}_2$. $\square$

Chapter 5

Bibliography

Agastya, M., F. Menezes, and K. Sengupta (2007): “Cheap talk, efficiency and egalitarian cost sharing in joint projects,” *Games and Economic Behavior*, 60, 1–19.

Aghion, P., P. Bolton, C. Harris, and B. Jullien (1991): “Optimal learning by experimentation,” *The Review of Economic Studies*, 58, 621–654.

Akerlof, G. A. (1970): “The market for" lemons": Quality uncertainty and the market mechanism,” *The Quarterly Journal of Economics*, 488–500.

Banks, J. S. and R. L. Calvert (1992): “A battle-of-the-sexes game with incomplete information,” *Games and Economic Behavior*, 4, 347–372.

Bar-Isaac, H. and J. Shapiro (2013): “Ratings quality over the business cycle,” *Journal of Financial Economics*, 108, 62–78.

Bergemann, D. and J. Valimaki (2008): “Bandit problems,” The New Palgrave Dictionary of Economics, 2nd ed. Macmillan Press.

Bolton, P., X. Freixas, and J. Shapiro (2012): “The credit ratings game,” *The Journal of Finance*, 67, 85–111.

Bolton, P. and C. Harris (1999): “Strategic experimentation,” *Econometrica*, 67, 349–374.

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.

Bouvard, M. and R. Levy (2009): “Humouring both parties: a model of two-sided reputation,” *working paper*.

BROCAS, I. AND J. D. CARRILLO (2007): “Influence through ignorance,” *The RAND Journal of Economics*, 38, 931–947.

COMPTE, O. AND P. JEHIEL (2009): “Veto constraint in mechanism design: inefficiency with correlated types,” *American Economic Journal: Microeconomics*, 1, 182–206.

CRAWFORD, V. P. AND J. SOBEL (1982): “Strategic information transmission,” *Econometrica*, 50, 1431–1451.

CREMER, J. AND R. P. MCLEAN (1988): “Full extraction of the surplus in Bayesian and dominant strategy auctions,” *Econometrica*, 1247–1257.

DOHERTY, N. A., A. V. KARTASHEVA, AND R. D. PHILLIPS (2012): “Information effect of entry into credit ratings market: The case of insurers’ ratings,” *Journal of Financial Economics*, 106, 308–330.

FARRELL, J. V. R. (1983): “Communication in games I: Mechanism design without a mediator,” *MIT working paper*, 334.

FELGENHAUER, M. AND E. SCHULTE (2013): “Strategic Private Experimentation,” *American Economic Journal: Microeconomics*, accepted for publication.

FORGES, F. (1990): “Universal mechanisms,” *Econometrica*, 58, 1341–1364.

——— (1994): “Posterior efficiency,” *Games and Economic Behavior*, 6, 238–261.

——— (1999): “Ex post individually rational trading mechanisms,” in *Current Trends in Economics*, Springer, 157–175.

FORGES, F. AND P. VIDA (2013): “Implementation of communication equilibria by correlated cheap talk: The two-player case,” *Theoretical Economics*, 8, 95–123.

GENTZKOW, M. AND E. KAMENICA (2012): “Disclosure of endogenous information,” *University of Chicago mimeo.*

——— (2014): “Costly Persuasion,” *The American Economic Review*, 104, 457–462.

GERARDI, D. (2004): “Unmediated communication in games with complete and incomplete information,” *Journal of Economic Theory*, 114, 104–131.

Glazer, J. and A. Rubinstein (2001): "Debates and decisions: On a rationale of argumentation rules," *Games and Economic Behavior*, 36, 158–173.

——— (2004): "On optimal rules of persuasion," *Econometrica*, 72, 1715–1736.

——— (2006): "A study in the pragmatics of persuasion: a game theoretical approach," *Theoretical Economics*, 1, 395–410.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144, 1397–1420.

Gresik, T. A. (1991): "Ex ante efficient, ex post individually rational trade," *Journal of Economic Theory*, 53, 131–145.

Henry, E. (2009): "Strategic Disclosure of Research Results: The Cost of Proving Your Honesty," *The Economic Journal*, 119, 1036–1064.

Hopenhayn, H. A. and F. Squintani (2011): "Preemption games with private information," *The Review of Economic Studies*, 78, 667–692.

Hörner, J., M. Morelli, and F. Squintani (2011): "Mediation and Peace. Cowles Foundation," *Unpublished manuscript.*

Hull, J. (2009): *Options, futures and other derivatives*, Pearson education.

Jackson, M. and H. Moulin (1992): "Implementing a public project and distributing its cost," *Journal of Economic Theory*, 57, 125–140.

Jovanovic, B. (1982): "Truthful disclosure of information," *The Bell Journal of Economics*, 36–44.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic experimentation with exponential bandits," *Econometrica*, 73, 39–68.

Lizzeri, A. (1999): "Information revelation and certification intermediaries," *The RAND Journal of Economics*, 30, 214–231.

MAILATH, G. J. AND A. POSTLEWAITE (1990): "Asymmetric information bargaining problems with many agents," *The Review of Economic Studies*, 57, 351–367.

MARIANO, B. (2008): "Do reputational concerns lead to reliable ratings?" *Financial Markets Group Discussion Paper*, 613.

MATHIS, J., J. MCANDREWS, AND J. ROCHET (2009): "Rating the raters: are reputation concerns powerful enough to discipline rating agencies?" *Journal of Monetary Economics*, 56, 657–674.

MATTHEWS, S. A. AND A. POSTLEWAITE (1989): "Pre-play communication in two-person sealed-bid double auctions," *Journal of Economic Theory*, 48, 238–263.

MCAFEE, R. P. AND P. J. RENY (1992): "Correlated information and mecanism design," *Econometrica*, 395–421.

MILGROM, P. AND J. ROBERTS (1986): "Relying on the information of interested parties," *The RAND Journal of Economics*, 18–32.

MYERSON, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics*, 10, 67–81.

——— (1985): "Bayesian equilibrium and incentive-compatibility: An introduction," in *Social goals and social organization*, ed. by L. Hurwicz, D. Schmeidler, and H. Sonnenschein, Cambridge Univ. Press, Cambridge, 229–260.

MYERSON, R. B. AND M. A. SATTERTHWAITE (1983): "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, 29, 265–281.

OTTAVIANI, M. AND P. SØRENSEN (2006): "Reputational cheap talk," *The RAND Journal of Economics*, 37, 155–175.

ROSENBERG, D., E. SOLAN, AND N. VIEILLE (2007): "Social Learning in One-Arm Bandit Problems," *Econometrica*, 75, 1591–1611.

ROTHSCHILD, M. (1974): "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 9, 185–202.

STRAUSZ, R. (2005): "Honest certification and the threat of capture," *International Journal of Industrial Organization*, 23, 45–62.

Williams, S. R. (1987): "Efficient performance in two agent bargaining," *Journal of Economic Theory*, 41, 154–172.

Curriculum Vitae

2008 - 2014	University of Mannheim, Germany Graduate School of Economics and Social Science PhD Studies in Economics
2012 - 2013	Yale University, USA Visitor Cowles Foundation
2003 - 2008	University of Mannheim, Germany Diplom Volkswirt
2006 - 2007	University of Paisley, Scotland B.Sc. in Business