Essays in Mechanism Design

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Benjamin Balzer

im Frühjahrs-/Sommersemester 2016

Abteilungssprecher: Prof. Dr. Carsten Trenkler,
Referent: Prof. Dr. Thomas Tröger,
Korreferent: Prof. Emanuele Tarantino, Ph.D.
Verteidigung: 29.06.2016

Curriculum Vitae

09/2010-06/2016	Ph.D. in Economics, University of Mannheim
09/2010-08/2012	M.Sc. in Economics, University of Mannheim
08/2011-06/2012	Yale University (*Visiting Ph.D. student*)
09/2007-07/2010	B.Sc. in Economics, University of Mannheim
10/2006-08/2007	Zivildienst, Bodelschwingh-Schule Kindergarten Sindelfingen
09/1997-06/2006	Abitur, Max Planck Gymnasium Böblingen
08/1995-07/1997	Deutsche Schule London
08/1993-07/1995	Sprachheischule Sindelfingen

Acknowledgements

I would like to thank my supervisor Thomas Tröger for his encouragement, support, valuable feedback and for sharing his very insightful view about mechanism design with me.

I also would like to thank my fellow graduate students and the Microeconomics Group for their valuable feedback and support, especially Emanuele Tarantino and Volker Nocke.

Moreover, I would like to thank Felix Jarman, Vincent Meisner and my co-author Johannes Schneider for their support and for pleasant and interesting company in academia and life.

Finally, I would like to thank my friends from outside academia and my family, Barbara, Catherina, Julia and Roland for their support.

Contents

Für Lutz Balzer

Chapter 1

Introduction

A central theme in economics is the asymmetric information between players, or market-participants. The theory of mechanism design provides powerful tools to resolve this information asymmetry, in order to achieve an improvement relative to the status quo. While resolving information asymmetry is often possible, it is typically costly (see e.g. Myerson and Satterthwaite, 1983). A players willingness to participate in the mechanism depends on his outside option, that is, on the situation that prevails if he refuses to participate. Often these outside options are modeled as exogenously given. As a consequence, the mechanism is designed in isolation and independent of the characteristics of the interaction, prevailing if the mechanism does not come into existence.

Common to the articles of this thesis is the emphasize on the less well-studied channel of endogenous outside options: In chapter 2 we consider the problem of a cartel, colluding prior to an auction. If a bidder does not participate in the cartel's mechanism, the non-cooperative play of the auction determines his payoff. In chapter 3 we solve the problem about how to best design an alternative dispute resolution mechanism that minimizes the number of cases entering the formal litigation process. If a player does not participate in the mechanism, litigation is played non-cooperatively.

Key channel in both chapters is the interaction between the default game and the characteristics of the mechanism. The play of the default game is influenced by the mechanism, because the former's characteristics determine the players inference about each other if the default game prevails. In turn, the mechanism is shaped by the characteristics of the default game, because these characteristics impact the player's process of inference if the mechanism fails to resolve the information asymmetry completely.

Collusion in Auctions

Chapter 2 consists of two articles. The first one analyzes bidder collusion in a second-price auction (hereafter SPA) and the second article analyses bidder collusion in a first-price auction (henceforth FPA). Absent of collusion, an auction creates competition between buyers, to elicit their private information, i.e. their valuation for the good that is for sale. However, the effectiveness of auctions threatened by bidders collusion. Before the auction takes place, bidders organize in cartels to suppress competition.

The main novelty of the articles contained in chapter 2 is that an insider proposes the collusive mechanism. So far, the literature has typically looked at models in which collusion is organized by an outsider, not a member of the cartel. In such models there is no strategic interaction between the cartel members present. The outsider proposes some efficient mechanism, i.e. the cartel member with the highest valuation receives the good.

In our model an insider organizes collusion and we solve a non-standard informed-principal problem. These problems differ from standard mechanism-design problems in that they do not answer normative questions, i.e. how to best designs the cartel's collusive mechanism. In contrast, informed-principal problems take a positive standpoint and answer the question what kind of mechanisms can arise in equilibrium as a choice of privately-informed players.

We find that efficient collusion cannot always be achieved in the SPA and can never be achieved in the FPA.

If an uninformed outsider organizes collusion, the cartel members do not learn anything from the proposed terms, so efficient collusion arises both in the SPA and FPA. Instead, an informed insider that designs collusion has an incentive to strategically reveal part of his private information via the collusive mechanism, and in this way gain a stronger bargaining position. Therefore efficient collusion cannot always be achieved in our model. We also find that the resulting strategic interaction depends on the auction format. The agent's bargaining position is determined by the non-cooperative play of the auction, following the agent's rejection of the collusive mechanism. As the non-cooperative play of the SPA is in weakly dominant strategies, the principal's mechanism proposal has no impact on this play. In contrast, in the FPA the principal designs the collusive mechanism to rationalize beliefs that punish the agents rejection.

Our results suggest that if the seller is only interested in welfare, like for example a government holding a procurement auction, he should propose a SPA. On the other hand, collusion in the SPA it is more profitable for the cartel than collusion in the FPA. Outside the model, if there are costs of being detected while colluding, the FPA should be less suspicious to collusion than the SPA.

Managing a Conflict

Chapter 3 is joint work with Johannes Schneider. This normative paper answers the question how to best design an alternative dispute resolution mechanism that minimizes the number of cases entering the formal litigation process, while at the same time keeping the rule of law in place.

Alternative Dispute Resolution (henceforth ADR) takes place before two litigants face each other in court. Litigation is an all-pay contest between players who are privately informed about their cost of evidence production, i.e. the quality of their cases. The player that produces the most evidence wins the case and receives a prize.

The main novelty of this paper is to consider a hybrid model between pure mediation, where the mechanism cannot enforce any allocation, and mechanism design, where full enforcement is feasible: An ADR mechanism can only propose and commit to settlement-shares, but cannot screen the player's private information directly, nor can dictate the player's behavior if litigation prevails. To screen the player's information, the mechanism has to rely on the non-cooperative litigation play. To assure the players voluntary participation in the ADR mechanism, full settlement is not always achievable. Thus, an optimal ADR mechanism breaks-down on the equilibrium path.

"Belief Management" plays the key role in the optimal mechanism: the way litigation is played after the break-down of ADR crucially depends on the information the players can infer from the breakdown of ADR, given the commonly-known features of the ADR mechanism. We find that the optimal mechanism induces beliefs such that the post-ADR litigation play is less cost-intensive than litigation play without ADR. As a consequence, litigation can be avoided often, while at the same time players are still willing to participate in the mechanism.

Besides highlighting the importance of belief management, we derive characteristics an optimal ADR mechanism satisfies: it gives rise to asymmetric breakdown beliefs, i.e. one player thinks his opponent has a good case, whereas the reverse is true for the other player. Moreover, an optimal mechanism is such that if full-settlement is not possible for all case combinations, then it does not promise full-settlement for any case combination, i.e. even if both players have bad cases they face each other in court with positive probability. These two conditions make the post-ADR litigation play relatively less cost-intensive. Finally, to assure truth telling in the ADR process, whatever a player infers from the breakdown of ADR is independent from his behavior during ADR, or in technical terms, a player with a good case holds the same beliefs about his opponent as a player with a bad case. These features can guide practitioners.

Chapter 2

Signaling and Inefficient Collusion in Auctions

2.1 Introduction

Auctions are pervasive allocation mechanisms. English auctions and sealed-bid first-price auctions are commonly used both in the public and private sectors to allocate and procure goods with limited supply and services among bidders who are privately informed about their valuations.[1] Collusion in auctions is a well-documented phenomenon: Before the auction takes place, potential bidders form a cartel and organize their joint bidding behavior to limit competition.[2] How does information asymmetry affect the strategic interaction between the cartel members? How does the auction format influence the cartel members' strategic interaction?

Key insights from the theoretical bidder-collusion literature suggest that the cartel can resolve the information asymmetry between its members at no cost, independently of the auction format. By employing an efficient mechanism that allots the member with the highest valuation the right to be the cartel's only bidder in the auction, the seller's rent is entirely absorbed. Yet, this literature has not analyzed how the bidders design the collusive agreement. Rather it generally assumes that a third party, outside the cartel, proposes the mechanism (see, for example, Graham and Marshall, 1987, Mailath and Zemsky, 1991, McAfee and McMillan, 1992, and Marshall and Marx, 2007).

In this paper, collusion is organized by cartel members, an assumption that squares with prevalent empirical evidence.[3] In the model, a member of the cartel (the principal) proposes the collusive side mechanism to the other cartel member (the agent), and all are privately informed about their types. The interaction between the cartel members boils down to a signaling game, and the collusive side mechanism is the signaling device. Its content can reveal part of the principal's private information,

[1]See, e.g., Marshall, Marx, and Meurer (2014) and Bichler et al. (2006).

[2]See, e.g., Marshall, Marx, and Meurer (2014) and Marshall and Marx, 2007 for various antitrust cases involving cartels that met prior to the auction.

[3]For example, in April 2008 the British Office of Fair Trading (OFT) closed one of its largest Competition Act investigations. 103 construction firms were found to have colluded with competitors on building contracts. In the course of this investigation, attention was drawn to the so-called "Calorie Club", consisting of rival builders frequently meeting and bargaining over the terms of collusion (e.g., see http://www.theguardian.com/business/2008/may/01/construction). More cases can be found in the OFT's decision, pages 395-425, accessible at http://webarchive.nationalarchives.gov.uk/20140402142426/http:/www.oft.gov.uk/OFTwork/competition-act-and-cartels/ca98/decisions/bid_rigging_construction.

which influences both the play of the side mechanism and, depending on the auction format, the auction's non-cooperative play in case collusion breaks down. By this, the auction format affects the agent's process of inference of the principal's type from the principal's proposal, and thus has an impact on the interaction in the cartel. The main contribution of this paper is to solve such a non-standard informed principal problem.

We derive and compare the implications of the auction format on the characteristics of bidder collusion. In particular, we are interested in whether the cartel can achieve its joint profit-maximizing benchmark of efficient collusion when it is organized by an insider. If collusion becomes less profitable then it is less likely to occur (see e.g., Tan and Yilankaya, 2007). The welfare implications of our analysis can guide the antitrust agency's decision about how much monitoring effort to exert on different auction formats (see Lopomo, Marshall, and Marx, 2005).

We take the following modeling approach: A passive seller auctions off an indivisible good with some public reserve price, either at a first-price auction (hereafter FPA) or a second-price auction (hereafter SPA).[4] Before the auction begins, the two potential bidders privately learn their type, i.e., their private valuation for the good, drawn by nature from commonly-known and independent distributions. Afterwards, the bidders might collude against the seller. Bidder P, the principal, proposes bidder A, the agent, a collusive side mechanism. Upon acceptance, the play of the side mechanism determines the enforceable[5] collusive allocation, consisting of side transfers between the two cartel members and a policy allotting the right to be the only bidder in the auction (henceforth the right). Upon rejection, the auction is played non-cooperatively.

Different types of principal prefer different collusive allocations. Thus, the principal's proposal potentially reveals part of his private information, influencing the agent's acceptance decision. This signaling possibility restricts the set of collusive allocations that can be supported in a perfect Bayesian equilibrium.

Our main result is that collusion is in general inefficient, both in the SPA and FPA.

Inefficiencies in the SPA are entirely driven by the principal's signaling motive. If he could commit himself ex-ante, before learning his type, to a side mechanism, collusion would be efficient: By proposing a mechanism that replicates the SPA's efficient non-cooperative play in weakly dominant strategies, the principal can fully extract the seller's rent.[6] That is, independently of his type, the principal obtains a constant payment equal to the expected profits that the seller would receive if the auction were to be played non-cooperatively.

[4] As the SPA is strategically equivalent to the English auctions, all our results concerning collusion in the SPA also apply to collusion in an English auction.

[5] The cartel's possibility to enforce the collusive agreement might be due to repeated interaction between the two potential bidders, either in future auctions or, if the bidders are firms, in some other market (see e.g., McAfee and McMillan, 1992). Given this, we essentially assume that the cartel possesses short-term commitment power, but long-term commitment is not feasible. When the environment (i.e., the distributions of the bidders' valuations) changes over time - for example, capacity constraints of firms colluding in a procurement auction - short-term commitment is a much less demanding requirement than long-term commitment. Whereas short-term commitment requires a collusive agreement only to be conditional on the commonly-known environment, long-term commitment requires the arrangement to be contingent on every possible future evolution of the environment.

[6] See Mailath and Zemsky, 1991, in which such mechanisms are presented.

However, if the principal proposes the mechanism after having privately learned his type, the efficient ex-ante optimal allocation cannot always be supported in equilibrium. Suppose there is a pooling equilibrium in which every principal type proposes a mechanism implementing this allocation. The principal's high types may have the incentive to deviate from this candidate equilibrium. Since the agent's outside option decreases in the principal's type, by using the collusive side mechanism as a signaling device, a high-type principal can increase its collusive rents. Essentially, he can credibly signal his type by proposing a mechanism that implements him the right with larger probability. Hence, signaling implies that rents are shifted from low to high principal types.

Rent shifting generates inefficiency. For the principal to receive the right with larger probability than in the efficient pooling equilibrium, the agent has to be excluded relatively more often from receiving the right. This reduces the joint collusive profits. Then, a high-type principal faces a trade-off. We show that the principal's decision to deviate from the efficient pooling equilibrium ultimately depends on the distribution of bidders' types.

If valuations are drawn from asymmetric distributions, with the principal having sufficiently large probability mass on low valuations, the signaling motive prevails and efficient collusion cannot be supported in equilibrium. Conversely, if the agent and the principal are close to being symmetric, or the principal is the stronger bidder, then the ex-ante optimal allocation is achievable under a pooling equilibrium. However, we show that only inefficient equilibria survive the Intuitive Criterion (Cho and Kreps, 1987).

The property that the agent's outside option decreases in the type of the principal also arises in the FPA. However, the non-cooperative play of the FPA depends on the commonly-known belief that the agent holds about the principal, and vice versa. The stronger a bidder thinks his opponent is, the less he will shade his bid below his valuation. As a consequence, the play of the side mechanism that replicates the non-cooperative play of the auction leads to different collusive allocations for different beliefs that the agent holds about the principal. If the agent thinks he faces a high-type principal, he is willing to pay the principal a high price for the right. Conversely, if the agent thinks the principal is a low type, then he will submit a low bid and thus expects to lose the auction. This allows those high-type principals to receive the right at no cost.

As a consequence, the high-type principals receive significantly larger payoff than low types by implementing the mechanism that replicates the non-cooperative play of the FPA. Since low-type principals will mimic high types, the collusive replication of the non-cooperative FPA allocation arises in equilibrium.

Yet, this does not imply that collusion is efficient. Indeed, if the agent rejects the collusive side mechanism proposed by the principal under this pooling equilibrium, the principal forms beliefs about the agent.

Being a strategic player, the principal maximizes his payoffs by proposing a side mechanism that leaves the highest agent type indifferent between accepting and rejecting the mechanism. Since rejecting the proposal would lead to an aggressive bidding behavior by the principal, the agent is willing to agree to collusive allocations that leave him with a rather low payoff.

Hence, in equilibrium, the collusive side mechanism replicates the non-cooperative play of an asymmetric FPA and thereby induces a fairly inefficient allocation. Collusion in the FPA is therefore always inefficient. Moreover, we show that it is more inefficient than the least inefficient equilibrium of the SPA. However, if one is only willing to select equilibria that survive the Intuitive Criterion, there is no general welfare ranking between the considered auction formats.

Our paper highlights that collusion is in general inefficient and the auction format has an impact on the cartels' interaction. Both findings stand in contrast to the common thought that cartels collude efficiently both in the FPA and the SPA.

Relation to the Bidding-Ring Literature Important contributions of the theoretical bidder-collusion literature model the cartel's interaction as being organized by an uninformed third party (see Graham and Marshall, 1987, McAfee and McMillan, 1992, Mailath and Zemsky, 1991 and Marshall and Marx, 2007). By this the cartel, often termed bidding ring, can achieve its joint profit-maximizing benchmark of efficient collusion. While this assumption is a realistic description of collusion in some cases, there is plenty of evidence that cartels are organized by insiders.[7] Rather, this assumption is typically employed to derive an upper bound of what a cartel can achieve despite the information asymmetry among its members (see e.g., Mailath and Zemsky, 1991). In our point of view, this approach overly abstracts from the interaction inside of the cartel. It does not take into account that the cartel members have to coordinate on a mechanism, *given* the information asymmetry.[8]

The way we introduce the resulting strategic interaction inside the cartel allows us to relate our findings to those of the literature on bidding rings. Indeed, we show that if the cartel operates in a SPA, the upper bound of efficient collusion is robust to a minimal form of signaling. However, for a cartel operating in the FPA, we find that the upper bound of efficient collusion is never achievable.

Although all our results are proven for two-bidder cartels, we argue in section 2.6 that the channels identified in the two-bidder case survive when increasing the number of cartel members.

[7]For example, in the antitrust case US v. Metropolitan Enterprises, Inc., 728 F.2d 444 (1984) (US v. Metropolitan); prior to a procurement auction held by the Oklahoma Department of Transportation, a manager of a highway paving construction firm proactively contacted potential competitors to arrange a meeting. The purpose of this meeting was to bargain over the right to become the cartel's only serious bidder in the procurement auction. More antitrust cases can be found at: http://www.justice.gov/atr/antitrust-case-filings-alpha. Most of them involve bid rigging. In many of them, the cartel members bargained over the terms of collusion without the help of an outsider.

[8]Asker, 2010 documents the case of a ring that repeatedly operated at English auctions for stamps. Throughout its existence, the ring used a pre-knockout auction as a revelation game to organize collusion. That is, prior to the target auction, each member submitted a sealed bid to a center, which then stayed active at the actual target auction until the bid reached the highest pre-knockout bid. Although allowing for side transfers between the ring members, such a knockout auction is not ex-post incentive-compatible. As the members' valuations and the beliefs about these valuations naturally changed over time, the ring members submitted too high knockout-bids. As a consequence, the ring frequently bought stamps at a price larger than the valuation of any ring member. The point we address is the following: even a ring that repeatedly colludes faces the necessity to organize collusion for each auction anew. That is, for each new auction, the members of the ring have to coordinate on a new revelation game. Clearly, we assume that the cartel in our model does so (see footnote 5 on page 6).

Relation to the Insider-Organized Bidder-Collusion Literature Eső and Schummer, 2004 introduce the signaling motive to the bidder-collusion literature. The cartel's side mechanism is a so-called take-it-or-leave-it bribing scheme. One of two bidders, say bidder P, both privately informed about their types, is able to submit a bribe to the other potential participant in a SPA, bidder A. By accepting the bribe, bidder A commits himself to stay out of the SPA.[9] In equilibrium, bidder A sometimes accepts bidder P's bribe offer, although the former has larger valuation than the latter, and thus collusion is inefficient. Rachmilevitch, 2013 analyzes the take-it-or-leave-it bribing scheme in the FPA. He finds that collusion does not occur at all - i.e., bidder P never bribes bidder A.

These models do not only introduce the signaling motive, but, in comparison to the literature on bidding rings, impose also harsh restrictions on the cartel's collusion technology. By bribing bidder A, bidder P can only buy the right to be the sole bidder in the auction. However, he cannot sell the right to bidder A or propose a mechanism that allots the right. Even if there was no information asymmetry between the bidders, bidder P would buy the right from bidder A types having larger valuation than himself, implying inefficient collusion. Indeed, restricting the cartel's side mechanism to the take-it-or-leave-it bribing scheme implies that the cartel's joint profit-maximizing benchmark of efficient collusion is excluded as possible outcome of the collusive interaction. Collusion can only be efficient, if it does not occur at all, that is, if the auction is played non-cooperatively for any realized type profile.

We build on Eső and Schummer, 2004 and Rachmilevitch, 2013 by allowing for a richer set of side mechanisms. In the model, we allow bidder P to propose one side mechanism. In particular, he can propose the take-it-or-leave-it bribing scheme side mechanism. However, we find that bidder P proposes a side mechanism that resolves the information asymmetry inside the cartel without relying on the non-cooperative play of the auction. Collusion therefore never breaks down, i.e., the auction is always played collusively. This implies that the equilibria of Eső and Schummer, 2004 and Rachmilevitch, 2013 do not survive beyond the case of the take-it-or-leave-it bribing scheme. Qualitatively, our findings stand in contrast to the papers mentioned above. Collusion in the SPA might be efficient, whereas collusion in the FPA always leads to inefficiencies.

Relation to the Informed-Principal Literature We build upon the methodologies developed by Myerson, 1983, Maskin and Tirole, 1992, and Mylovanov and Tröger, 2015 to solve an informed-principal problem. Our paper is one of only a few instances that solves an informed-principal problem with two-sided private information under common values that emerge because the agent's outside option is determined by the play of a default game.[10] Moreover, to the best of our knowledge, our paper is the first one that solves an informed-principal problem in which

[9] Francetich and Troyan, 2012 extend this setup to a common value environment. In their model, an informed principal not only designs the collusive side mechanism but also runs it. This lack of principal commitment reduces the set side mechanisms to the take-it-or-leave-it bribing scheme.

[10] To the best of our knowledge, Quesada, 2004 and Che and Kim, 2006 are the only papers that solve a similar problem. However, in their collusion setups, the collusion-proof principle applies and in turn implies that the RSW (Rothschild-Stieglitz-Wilson) allocation is the unique equilibrium allocation. This allocation is degenerate in the sense that it is a replication of the non-cooperative play of the default game.

the agent's outside option is determined by a non-trivial, belief-dependent game, the FPA.

Myerson, 1983 formulates the inscrutability principle, which states that it is without loss of generality to focus on pooling equilibria of the mechanism-selection game.[11] Moreover, he introduces the concept of safe allocations. An allocation is safe if it is incentive-compatible and individually rational for any belief about the principal having point mass. Maskin and Tirole, 1992 introduce the RSW (Rothschild-Stieglitz-Wilson) allocation. This allocation is defined as the best safe allocation from every principal type's perspective. The following result is established: Suppose there exists some *full-support* belief about the principal such that the RSW allocation is undominated from the principal's perspective. Then every allocation which leaves all principal types with a weakly larger payoff than the RSW allocation is an equilibrium allocation.

Mylovanov and Tröger, 2015 consider a general framework which applies to signaling games where the sender's message space is sufficiently rich. In particular, they introduce the secured-payoff vector. Intuitively, this vector consists of the max-min payoff, where the max is with respect to the principal's mechanism proposal and the min with respect to the belief about the principal, each principal type receives in any (continuation) equilibrium starting after his mechanism proposal. They establish the following result: If there exists *some* belief about the principal such that the secured-payoff vector is undominated from the principal's point of view, every allocation that leaves all principal types with a weakly larger payoff than their secured payoff is an equilibrium allocation.

Focusing on the SPA, we show the secured-payoff vector is implemented by the RSW allocation. This is an intuitive result, given the technical similarity between a problem in which the agent's outside option is determined by a game in dominant strategies and a problem where the outside option is fixed but type-dependent, as in Maskin and Tirole, 1992. In addition, we make use of an insight from the latter paper while characterizing the RSW allocation; instead of solving for the principal types' most preferred safe allocation jointly, one can equivalently find the RSW allocation by solving a sequence of linked maximization problems, one for each principal type. Having characterized the optimal solution of these problems, we verify the existence of a belief about the principal such that the RSW allocation is undominated. Because this belief does not in general have full support, we build the existence of equilibrium by drawing upon Mylovanov and Tröger, 2015.

Focusing on the FPA, we show that the secured-payoff vector is *not* implemented by the RSW allocation, and the latter allocation cannot be supported in equilibrium. We establish the existence of a certain side mechanism, say $\tilde{m}$, that is accepted for any belief the agent holds about the principal. Depending on this belief, the play of $\tilde{m}$ implements different allocations. That is, $\tilde{m}$ is not a direct revelation mechanism and does not implement a safe allocation. We show that the secured payoff is implemented by the play of $\tilde{m}$ for some belief about the principal. Moreover, we verify that the allocation being implemented by the play of $\tilde{m}$, with the agent having prior beliefs about the principal, is the unique equilibrium allocation.

[11] By the revelation principle, the play of a separating equilibrium of the mechanism-selection game induces some incentive-compatible allocation. Because the principal is a player of the game he proposes, this allocation can directly be implemented in a pooling equilibrium. The principal offers a direct revelation mechanism that implements the corresponding allocation.

Our analysis provides useful insights to advance our knowledge of the solution to mechanism-design problems with an informed principal in quasilinear, independent-private-values environments under the following context (i) in case outside options are non-trivial, (ii) in case the types of the principal can be screened - that is, the principal's incentive constraints are non-trivial, and (iii) to characterize allocations that survive the Intuitive Criterion.

Concerning (i), the result that the RSW allocation is not an equilibrium allocation if collusion is in the FPA is both surprising and important, as it indicates that this is a general feature in informed-principal problems in settings where outside options are determined by the play of a non-trivial game. Concerning (ii), the ideas employed to provide conditions assuring that the principal's ex-ante optimal allocation can be supported in equilibrium of the game might be useful for literature interested in the question of the irrelevance of the principal's private information (see Koessler and Skreta, 2014 and the references therein).[12] Concerning (iii), we characterize the set of equilibrium allocations that survive the Intuitive Criterion. Maskin and Tirole, 1992 provide a characterization for the case of one-sided private information under a sorting assumption. They find that only the RSW allocation survives the refinement. If the sorting assumption holds in our two-sided private information setting, we arrive at the same result. However, if the sorting assumption fails, the RSW allocation is not the unique allocation that survives the Intuitive Criterion.

Other related Literature We contribute to a strand of the literature initiated by Robinson, 1985 that argues that the FPA is less suspicious to collusion than the SPA. In order to gain from the right to be a cartel's only bidder in the FPA, a member has to submit a low winning bid at the auction. This creates incentives for other cartel members to deviate from the collusive agreement and to enter the auction unexpectedly. The results of Marshall and Marx, 2007 and Lopomo, Marx, and Sun, 2011 suggest that collusion in a one-shot FPA can hardly be rationalized without assuming that the cartel can enforce the collusive agreement. However, collusion in FPAs is documented, and enforcement does not seem to be impossible in general. Indeed, there is a large literature modeling collusion explicitly as repeated interaction between the cartel members. For example, Aoyagi, 2003 and Aoyagi, 2007 show that due to grim trigger strategies, the cartel can enforce the terms of collusion independently of the auction format.

Our results highlight that the strategic interaction inside of a cartel operating at a FPA prevents it from reaching an efficient and profit-maximizing agreement, *although* the cartel is powerful enough to exchange side transfers and to enforce the agreement.

This paper is related to a small literature arguing for inefficient collusion in English auctions or in SPAs. In Lopomo, Marshall, and Marx, 2005, the members of a non-all-inclusive cartel are not able to communicate before an English auction

[12] If collusion is in the SPA, the principal's common-knowledge payoff (i.e., the payoff he would receive in case the agent knew his type) cannot be implemented in equilibrium. Yet, the principal's ex-ante optimal payoff might be implemented. Conversely, in the FPA, the principal's ex-ante optimal payoff cannot be implemented; however, he receives a larger payoff than his common-knowledge payoff. That is, in the SPA the principal weakly prefers the agent to know his type (from an ex-ante point of view), whereas in the FPA the principal strictly prefers the agent not to know his type, both ex-ante and ex-interim.

in order to decide on the maximal cartel bid, submitted by one of the members to purchase the good for a post-knockout auction among them. Instead, the cartel members must communicate via their bidding behavior. This bidding behavior implies an inefficient allocation of the good. In Garratt, Tröger, and Zheng, 2009, the possibility of resale leads to collusive bidding behavior in a second-price or English auction, implying inefficiencies in the final allocation of the good.

In our model the inefficiency results from the signaling channel for which pre-auction communication between the cartel members is crucial.

Finally, for the sake of clarity, we do not relate to the large literature on collusion-proof mechanisms, to which the (weak) collusion-proof principle - see Laffont and Martimort, 1997, Laffont and Martimort, 2000 - applies: The designer of the grand game can restrict attention to game forms, or mechanisms, that leave no cartel member better off by colluding. We depart from this literature by restricting the seller's mechanism to a SPA or FPA.

Outline We proceed by giving an illustrative example of some of our results. Afterwards, we introduce the model (section 2.2) and continue in section 2.3 by presenting our main results for collusion in the SPA. We proceed in section 2.4 by focusing on the case of collusion in the FPA. In section 2.5, we consider extensions for the case of collusion in the SPA. In particular, we refine the set of equilibria by the Intuitive Criterion in section 2.5.1 and consider the case of a cartel that cannot enforce the right to be the only bidder in the auction 2.5.3. In section 2.7, we conclude and discuss our results.

2.1.1 Illustrative Example

We begin with an example that illustrates the channels behind the main results of the paper. Suppose there are two bidders, indexed by $i \in \{A, P\}$. Both are interested in winning an indivisible good, which is auctioned off at a SPA with zero reserve price. Bidder A's type, j, is distributed uniformly on the interval $[0, 1]$ with distribution $F(\cdot)$. Assume for simplicity that bidder P's type, θ_k, can be low ($k = 0$) or high ($k = 1$). Independently of bidder A's valuation draw, each event occurs with probability β_k. Assume that $\theta_0 = 1/4 < \theta_1 = 5/8$. These primitives are common knowledge.

Before the auction begins, and after each bidder has privately learned the realization of his valuation, bidder P - the principal - proposes bidder A - the agent - a collusive side mechanism. If the agent accepts the mechanism, its play determines the collusive allocation that consists of a policy allotting the right to be the only bidder in the auction and transfer payments between both bidders. If the agent rejects the mechanism, both the agent and the principal bid their valuation in the auction. Because there is no reserve price, a sole bidder in the auction receives the good at price zero. Hence, each player's valuation of the right is equal to his valuation of the good. We want to examine whether or not efficient collusion can arise in equilibrium. That is, consider a hypothetical equilibrium in which the principal, independently of his type, proposes a direct revelation mechanism that allots the right for any realized type profile to the player with the highest valuation. If $q^i(\cdot)$ stands for bidder i's interim expected allotment probability, efficient collusion implies

$$q^P(k) = F(\theta_k), \qquad q^A(j) = \sum_{k=0}^{1} \mathbb{1}[j > \theta_k]\beta_k.$$

These allotment probabilities can be implemented by means of a linear-separable transfer scheme. If $h^i(\cdot)$ denotes bidder i's type-dependent payment, the following specification assures incentive compatibility:

$$h^P(0) = 0, \qquad h^P(1) = \theta_0[F(\theta_1) - F(\theta_0)], \qquad h^A(j) = \sum_{k=0}^{1} \mathbb{1}[j > \theta_k]\theta_k\beta_k.$$

In any undominated equilibrium, these type-dependent payments stay inside of the cartel. Bidder i pays $h^i(\cdot)$ to bidder $-i$, and in addition the principal asks the agent to pay a lump-sum transfer equal to $\mathbb{E}_k[h^P(k)]$. Since the allotment of the right equals the allotment of the good induced by the non-cooperative play of the auction, $h^A(\cdot)$ is optimally set - from the principal's perspective - equal to the expected payment the agent would have to pay if the auction were to be played non-cooperatively. If we denote the resulting allocation by ϕ^e, the principal's payoff, $U^P_{\phi^e}(\cdot)$, takes the form:

$$U^P_{\phi^e}(0) = F(\theta_0)\theta_0 + \mathbb{E}_{j,k}[h^P(k) + h^A(j)], \qquad U^P_{\phi^e}(1) = F(\theta_1)(\theta_1 - \theta_0) + U^P_{\phi^e}(0).$$

The above presented allocation maximizes the bidders' joint collusive profits and leaves every type of the agent with exactly his outside option. Thus, the principal can extract the entire profits, and efficient collusion is ex-ante optimal from the principal's point of view. This is true in general. Every ex-ante optimal allocation induces efficient collusion.[13]

However, the extracted profits, $U^P_{\phi^e}(0)$, are shared equally between both types of the principal. Because it is the high type that contributes the most to those profits, he might have an incentive to use the side mechanism as a signaling device.

For example, he can propose a mechanism that corresponds to the following game form, which is accepted by the agent for any off-path belief about the principal's type: The game begins with the principal's choice of one, out of a total of two, specified menus. Afterwards, the agent determines the allocation by picking an element of the offered menu. One might think of each of these menus as specifying a price, p_k, at which the agent can buy the right. If the latter player does not want to pay the price, he receives a compensation, c_k, for the duty to stay out of the auction. For the latter specified parameter $\tilde{j}$, with $\tilde{j} > \theta_1$, let these two menus take

[13]Since the principal has finitely many types, the allotment policy does not uniquely determine the payoff of the principal's types. Thus, in general, there are infinitely many ex-ante optimal allocations. Each such allocation features efficient collusion but differs concerning the distribution of the payoff between the principal's types due to the freedom concerning the principal's adjacent incentive constraints. In the subsequent analysis, we focus on the ex-ante optimal allocation in which every principal type's upward adjacent incentive constraint is satisfied with equality. This is the most preferred ex-ante optimal allocation of the largest principal type. Since efficient collusion can be supported in equilibrium if and only if the largest principal type prefers the ex-ante optimal allocation to the RSW allocation, this approach is without loss of generality.

the following form:

$$\{p_0 = \theta_0,\ c_0 = 0\}, \qquad \{p_1 = \theta_1,\ c_1 = \tilde{j} - \theta_1\}.$$

Suppose the agent accepts the game form and menu 0 is offered. In this instance, agent types below θ_0 pick the compensation, as the price for the right is weakly higher than their valuation ($j - p_0 = j - \theta_0 \leq 0 = c_0$). Agent types above θ_0 buy the right at price θ_0 and thus receive payoff $j - \theta_0 \geq 0 = c_0$. Because menu 0 induces an efficient allocation, every type of the agent receives the same payoff as from playing the SPA non-cooperatively against principal type 0.

If menu 1 is offered, agent types below $\tilde{j}$ pick the compensation as $j - p_1 = j - \theta_1 \leq \tilde{j} - \theta_1 = c_1$. By this, they receive payoff larger than from playing the auction non-cooperatively against principal type 1. Agent types above $\tilde{j}$ buy the right and are left with payoff $j - \theta_1 \geq \tilde{j} - \theta_1 = c_1$, equals the payoff from playing the auction non-cooperatively against principal type 1.

Since the above described behavior of the agent is independent of his belief about the principal, the agent accepts the game form if principal type k optimally proposes menu k.

If this is the case, the following allocation is induced (where $t^P(k)$ denotes the expected transfer payment of principal type k):

$$q^P(0) = F(\theta_0), \qquad q^P(1) = F(\tilde{j})$$

$$t^P(0) = -F(\theta_0)\theta_0, \qquad t^P(1) = -[1 - F(\tilde{j})]\theta_1 + F(\tilde{j})[\tilde{j} - \theta_1].$$

Indeed, principal type 0 is indifferent between proposing menu 0 and 1. If type 0 offers menu 0, he receives payoff $q^P(0)\theta_0 - t^P(0) = \theta_0$. Thereby he attains the upper bound of any payoff he could receive if the agent knew his type. Given menu 0, menu 1 is constructed by introducing the smallest inefficiencies necessary to assure that principal type 0 does not want to offer menu 1. That is, $\tilde{j}$ solves

$$q^P(1)\theta_0 - t^P(1) = \theta_0 \iff F(\tilde{j})\theta_0 - [1 - F(\tilde{j})]\theta_1 + F(\tilde{j})[\tilde{j} - \theta_1] = \theta_0.$$

The solution is such that $\tilde{j} = 3/4 > \theta_1$, and thus the allocation is inefficient in favor of the principal. Principal types differ in their valuation for the good. Hence, the screening parameter is the probability with which they receive the right. In order to credibly signal his type, principal type 1 has to offer an inefficient allocation in favor of him. In turn, agent types with higher valuation than θ_1 are necessarily excluded from receiving the right.

The described allocation is called RSW allocation. Let us denote by $U^P_{RSW}(1)$ the high principal type's payoff from this allocation. Given the specified valuations, he prefers to deviate from the hypothetical equilibrium by offering the RSW allocation if:

$$U^P_{RSW}(1) > U^P_{\phi^e}(1)$$

$$\iff F(\tilde{j})(\theta_1 - \theta_0) + \theta_0 = 17/32 > F(\theta_1)(\theta_1 - \theta_0) + U^P_{\phi^e}(0) = 27/64 + \beta_1(9/64).$$

Hence, if $\beta_1 < \frac{7}{9}$, efficient collusion cannot be supported in equilibrium. We verify in section 2.3 that an allocation can be supported in equilibrium if and only if every type of the principal receives a weakly larger payoff than his RSW payoff. As a consequence, if $\beta_1 \geq 7/9$, efficient collusion can be supported in equilibrium.

This example illustrates a general feature of collusion in the SPA. The high principal type faces a trade-off. On the one hand, efficient collusion is desirable because it allows the principal to entirely extract the cartel's maximized joint profits. On the other hand, using the mechanism as a signaling device and proposing the RSW allocation allows the principal to shift rents from type 0 to type 1. Yet, this rent shift necessarily induces inefficiencies, which decrease the rent to be shifted. In which direction the trade-off is resolved depends on β_1. The larger β_1, the larger the negative impact of the inefficiencies on the rent to be shifted. This feature holds in general. The stronger the principal is in stochastic terms, the less desirable it is to introduce inefficiencies.

Now suppose the auction format is a FPA. The non-cooperative play of the auction depends on the belief the agent holds about the principal, and vice versa. Suppose for the moment the FPA is played non-cooperatively with degenerate off-path belief concentrated on the highest agent type, $j = 1$, and some arbitrary belief about the principal. Let $\tilde{\beta}_k$ denote an arbitrary belief about principal type k.

Then there are, in particular, the following two possible types of equilibria. In any such, principal type 0 submits a bid equal to his type:

If the agent thinks the principal is fairly strong, that is, $(\theta_1-\theta_0)/(1-\theta_0) < \tilde{\beta}_1$, then both principal type 1 and the highest agent type, $j = 1$, submit bids equal to θ_1. In this instance, we assume that agent type $j = 1$ receives the right with certainty and is left with payoff $1 - \theta_1$. Any agent type below 1 either submits bid θ_1, θ_0, or 0.

If the agent thinks the principal is fairly weak, that is, $(\theta_1 - \theta_0)/(1 - \theta_0) \geq \tilde{\beta}_1$, principal type 1 mixes against agent type $j = 1$ on the interval $[\theta_0, 1\tilde{\beta}_1 + \tilde{\beta}_0\theta_0]$ and the latter is left with payoff $\tilde{\beta}_0(1 - \theta_0)$. Any agent type below 1 either submits a bid equal to θ_0 or 0.

Hence, the agent's ex-post outside option, given principal type k, not only depends on k but also on the belief the agent held about the principal at the point in time where he submitted his bid.

The principal can exploit this feature of the FPA by proposing the mechanism $\tilde{m}$, corresponding to the following game form.

There is a pool consisting of three menus:

$$\{p_\emptyset = 0, \quad c_\emptyset = 0\}, \qquad \{p_0 = 0, \quad c_0 = \theta_0\}, \qquad \{p_1 = 0, \quad c_1 = \theta_1\}.$$

The game begins with the agent selecting one menu. Afterwards, the principal picks an element from the chosen menu and thereby determines the allocation.

Each menu in the constructed pool consists of a price, p, at which the principal can buy the right, and a compensation, c. If the principal picks the compensation, he commits to stay out of the auction and the agent receives the right. If the agent proposes menu $\check{k}$, with $\check{k} \in \{0, 1\}$, principal type k picks the compensation if and only if $k \leq \check{k}$ (as $c_{\check{k}} = \theta_{\check{k}} \leq \theta_k - p_{\check{k}} = \theta_k$). If the agent proposes menu $\emptyset$, every principal type picks the right and the agent is left with payoff 0. Thus, for any $\tilde{\beta}$, selecting the optimal menu is from the agent's perspective exactly the same problem as deciding upon the optimal bid in the non-cooperative play of the FPA. Hence, independently of his belief about the principal, every agent type accepts $\tilde{m}$ and receives payoff equal to his expected outside option.

Now note that according to the play of $\tilde{m}$ principal type 0 receives, depending on $\tilde{\beta}$, at least payoff θ_0, whereas for any $\tilde{\beta}$, type 1 receives payoff θ_1. This is a fairly large payoff for principal type 1. In fact, it is equal to the largest payoff he could receive if his type were commonly known. By this, principal type 1 has no incentive to deviate from a hypothetical equilibrium in which $\tilde{m}$ is proposed in order to signal his true type to the agent.

Moreover, other equilibria do not exist. We show in section 2.4 that any collusive allocation that differs from the one implemented by the play of $\tilde{m}$ with prior beliefs about the principal, leaves principal type 1 with strictly lower payoff than θ_1. Given this, consider a hypothetical equilibrium that gives rise to a different allocation. Principal type 1 would propose $\tilde{m}$ and thereby secure himself the payoff θ_1.

As for intuition, by replicating the FPA, $\tilde{m}$ heavily exploits the bid-shading property of the FPA. If the agent thinks he faces the high principal type with large probability, the former submits a high bid and expects to win the auction for sure. In such a situation, the agent is willing to pay the principal a high price for the right. If the agent thinks he will face a low principal type, the former submits a low bid and expects to win the auction only against the low type of the latter. Thus, the agent is willing to let the high principal type win the auction. Hence, when proposing $\tilde{m}$, there essentially does not exist a worst (off-path) belief about the principal, from the high principal type's perspective, with which different equilibria could be supported.

In general, for any prior belief β about the principal, the allocation implemented by the play of $\tilde{m}$ is the unique (in terms of payoffs) equilibrium allocation and more inefficient than the least inefficient equilibrium allocation if collusion is in the SPA.

2.2 The Model

Primitives

We consider an independent-private-values setting with two risk-neutral bidders who are interested in consuming an indivisible good. The good is initially possessed by a seller who has no value for it.

The valuation of bidder A, the agent, is denoted by j, with $j \in [0, J] := \Theta^A$. j is distributed according to the distribution $F(\cdot)$, with continuous density $f(\cdot)$ on the interval $(0, J]$. $F(\cdot)$ may have a point mass at zero. We impose a regularity condition on $F(\cdot)$:

Assumption 1. $j + \frac{F(j)}{f(j)}$ *is non-decreasing in* j, $\forall j > 0$.

Bidder P, the principal, has finitely many possible valuations, indexed by k, with $k \in \{0, 1, 2..., K\} =: \Theta^P$.[14] The valuations are drawn from the ordered set $\{\theta_0, ..., \theta_K\}$, where we assume without loss of generality that $\theta_0 = 0$ and $\theta_1 > 0$.[15] Valuation θ_k realizes with probability β_k, where the cumulative probability function is given

[14] Since we employ recursive methods in the analysis, it is frequently the case that $k = -1$. In this instance, we understand terms which involve a quantity being indexed with -1 as zero. Similarly, whenever a term consists of a quantity being indexed by $k = K + 1$, this term is zero.

[15] The finite type space of the principal is assumed to avoid the technical difficulties of signaling games in which the sender has a continuous type space. However, our model allows arbitrarily close approximations of a continuous type space.

by $B(k) = \sum_{v=0}^{k} \beta_v$. Let $\beta := \{\beta_0, ..., \beta_K\}$ denote the (true) prior probabilities, and indicate by $\tilde{\beta} \in B_0 := \{b \in \mathbb{R}_+^K | \sum_{k=0}^{K} b_k = 1, b_k \geq 0\}$ an arbitrary belief of the agent about the principal's type. Furthermore, we define by $\tilde{B}(k) := \sum_{v=0}^{k} \tilde{\beta}_v$ the resulting cumulative probability function.

The type space is denoted by Θ, with $\Theta := \Theta^A \times \Theta^P$. We impose the following non-triviality assumption on Θ:

Assumption 2. $\theta_1 < J$.

That is, there is at least one type of the principal having strictly positive, and strictly lower valuation than the highest agent type.

Throughout most of the paper, we model the seller as passive. We assume that he holds either an English auction, second-price auction (hereafter SPA), or a first-price auction (hereafter FPA). Since the English auction is strategically equivalent to the SPA, we analyze the following format:

Second-Price Auction. *The seller announces a reserve price r, with $r < \theta_1$. Both bidders submit a sealed bid. The highest bid above the reserve price wins the single indivisible object. The winner pays the maximum between the second highest bid and the reserve price. If both bids are equal, the auctioneer assigns the good to bidder P.*[16]

The rules of the FPA are given by:

First-Price Auction. *The auctioneer announces a reserve price r, with $r < \theta_1$. Both bidders submit a sealed bid. The highest bid above the reserve price wins the single indivisible object. The winner pays his bid. If both bids are equal, the auctioneer assigns the good to the player with the highest valuation.*[17]

The bidders are able to collude in the auction. We model collusion as a three-period interaction, which begins after each bidder has privately learned his valuation. The collusive interaction, which we frequently denote as grand game, features the following sequence of events:

Timing.

- *t=0: Each bidder privately learns his valuation.*
- *t=1: The principal proposes to the agent a collusive side mechanism.*
- *t=2: The agent accepts or declines the mechanism. In the former instance, stage t=2′ applies. If he rejects the mechanism, go directly to t=3.*
- *t=2′: The proposed side mechanism is played and determines the enforceable terms of collusion.*
- *t=3: The auction is played, possibly under collusion. Whenever the terms of collusion allow a bidder to bid in the auction, or if the agent rejected the collusive side mechanism, the bidders play the auction non-cooperatively.*

[16] This tie-breaking rule is innocuous. We assume it to economize on notation.
[17] The efficient tie breaking rule is a standard trick in the literature to simplify the analysis of the FPA (e.g., see Bergemann, Brooks, and Morris (2013)).

We are interested in perfect Bayesian equilibria.

In period $t = 1$, bidder P proposes bidder A a collusive side mechanism, or game form (we use these terms interchangeably). Similar to Maskin and Tirole, 1992 we allow the principal to offer any simultaneous-move game form with finitely many actions for the principal and assume the existence of a public randomization device.[18] If the agent accepts the mechanism, both bidders simultaneously choose an action or message. The resulting collusive outcome is then determined by m.

A collusive outcome allows the bidders to (i) decide which of them is allowed to bid in the auction, and to (ii) exchange side payments in a budget-balanced way. Thus, by assumption, no bidder can be committed to other actions in the auction than staying out.[19] Therefore, if a player receives the right to participate as the only bidder in the auction, he uses it if and only if his valuation is above the reserve price. In most of the paper, we formally set the reserve price to zero and capture its effects by allowing for a point mass in the bidders' type distributions at zero.

A (collusive) allocation is a complete type-dependent description of the outcome that results from the bidders' interaction. A (collusive) allocation is defined as a mapping from the bidders' type space into the set of (collusive) outcomes:

$$\phi(\cdot) = (q_\phi(\cdot), q_\phi^{A,P}(\cdot), t_\phi(\cdot)) : \Theta \to [0,1]^2 \times [0,1] \times \mathbb{R}^2.$$

The allocation consists of a policy that allots the right to be the only bidder in the auction (henceforth the right), $q_\phi(\cdot) = (q_\phi^A(\cdot), q_\phi^P(\cdot)) : \Theta \to [0,1]^2$. The probability with which bidder i, $i \in \{A, P\}$, receives the right is denoted by $q_\phi^i(\cdot)$. Moreover, $q_\phi^{A,P}(\cdot) : \Theta \to [0,1]$ stands for the probability that both bidders are allowed to participate in the auction. Lastly, $t_\phi(\cdot) = (t_\phi^A(\cdot), t_\phi^P(\cdot)) : \Theta \to \mathbb{R}^2$ denotes the transfer scheme, with $t_\phi^i(\cdot)$ being the payment made by bidder i. An allocation satisfies resource feasibility, that is, for any type profile $(j,k) \in \Theta$ it is the case that

$$q_\phi^A(j,k) + q_\phi^P(j,k) + q_\phi^{A,P}(j,k) \leq 1.$$

A collusive allocation is budget balanced; that is, $\forall (j,k) \in \Theta$ it holds that:

$$t_\phi^A(j,k) + t_\phi^P(j,k) \geq 0.$$

We refer to this property with the symbol $(BB)_{j,k}$. Moreover, we impose the restriction that transfers are uniformly bounded.

[18] To be more precise, Maskin and Tirole, 1992 assume that the principal can only offer finite game forms. This is a sufficient condition to ensure the existence of an equilibrium of the continuation game starting after $t = 1$ with the agent's decision whether to accept or to reject the proposed game form. Because the principal has finitely many types, every relevant equilibrium allocation in our model can also be implemented by a game form with only finitely many actions for the agent. But, for convenience, we analyze the situation in which the game form can specify infinitely many actions for the agent and assume that a continuation equilibrium always exists. Given the above justification, this is a very mild assumption.

[19] This assumption implies that a bidder with valuation below the reserve price cannot be forced to submit a positive bid in the auction. If we dropped this assumption and analyzed collusion in the SPA, then, whenever there is a non-trivial reserve price, the RSW allocation would be dominated for all beliefs about the principal. By this, the analysis would largely be complicated. On the other hand, if there is no reserve price, then this assumption is without loss of generality.

Note, the definition of ϕ is general enough to incorporate cases in which the collusive interaction results in the non-cooperative play of the auction. That is, if agent type j rejects the side mechanism, then $q_\phi^{A,P}(j,\cdot) = 1$ and $t_\phi^A(j,\cdot) = t_\phi^P(j,\cdot) = 0$.

A collusive mechanism m is in the set of feasible game forms M if and only if any play of m results in a collusive allocation that satisfies the above conditions.

Given the general setting introduced above, we first consider collusion in the SPA.

2.3 Collusion in the SPA

Continuation Game after period 1 With the beginning of period $t = 2$, the agent decides whether to accept or to reject the mechanism proposed by the principal, m. At the start of this continuation game, the agent's belief about the principal can differ from the prior belief β. Suppose in the following that the agent holds some arbitrary belief $\tilde{\beta}$.

If the agent rejects m, the SPA is played non-cooperatively. We restrict our attention to the equilibrium in weakly-dominant strategies.

We denote by $U_D^A(j,k)$ the ex-post utility of agent type j when playing the SPA non-cooperatively against principal type k. That is, $U_D^A(j,k) := \max\{0, j - \theta_k\}$. $U_D^A(j,\tilde{\beta})$ denotes agent type j's interim expected default utility when holding the belief $\tilde{\beta}$ about the principal's type. That is, $U_D^A(j,\tilde{\beta}) := \sum_{k=0}^{K} \tilde{\beta}_k U_D^A(j,k)$. Let $t_D^A(j,\tilde{\beta})$ denote agent type j's expected transfer payment submitted to the seller. Let $q_D^A(j,\tilde{\beta})$ be agent type j's expected probability of winning the good via the non-cooperative play of the SPA. Similarly, we denote the respective objects for the principal holding prior beliefs about the agent as $U_D^P(j,k)$, $U_D^P(k)$, $t_D^P(k)$ and $q_D^P(k)$.

If the agent accepts the game form m, the play of the side game determines the collusive allocation ϕ. Given a type profile $(j,k) \in \Theta$, agent type j's ex-post payoff takes the form:

$$u^A(\phi(j,k), j, k) := j q_\phi^A(j,k) - t_\phi^A(j,k) + q_\phi^{A,P}(j,k) U_D^A(j,k).$$

Principal type k's ex-post payoff reads:

$$u^P(\phi(j,k), j, k) := \theta_k q_\phi^P(j,k) - t_\phi^P(j,k) + q_\phi^{A,P}(j,k) U_D^P(j,k).$$

In any continuation equilibrium, the agent accepts the proposed mechanism at the beginning of stage $t = 2$ if and only if the continuation value from accepting the mechanism is larger than the expected utility from playing the SPA non-cooperatively. Because this latter payoff depends on the type of the principal, the analyzed game is an informed-principal problem under common values.

Take any allocation induced by m, that is, through some equilibrium play of the continuation game starting after the proposal of m. Note, a collusive side mechanism can replicate any outcome that results from the bidders' interaction. In particular, it can replicate the outcome that results if the agent rejects the mechanism m by specifying transfers of zero and allowing both bidders to bid non-cooperatively in the SPA. Therefore, the revelation principle for Bayesian games implies that it is without loss of generality to focus attention on (continuation) equilibria, which are such that the principal proposes a collusive side mechanism that is unanimously accepted by all agent types. Thus, a collusive allocation can be induced through some (continuation) equilibrium play if and only if it can be implemented by a direct revelation mechanism. Fix a (collusive) allocation ϕ and suppose the agent submits report $\hat{j}$ and the principal reports his type truthfully. $q^A_\phi(\hat{j}, \tilde{\beta})$ denotes the agent's expected likelihood of receiving the right. $t^A_\phi(\hat{j}, \tilde{\beta})$ stands for the agent's expected transfer. Whenever $\tilde{\beta} = \beta$ - i.e., the agent has the prior belief about the principal - we suppress the argument β. We denote the respective objects for the principal holding prior beliefs about the agent as $q^P_\phi(\hat{k})$ and $t^P_\phi(\hat{k})$.

Agent type j's interim utility, given report $\hat{j}$, takes the form:

$$U^A_\phi(\hat{j}, j, \tilde{\beta}) := \sum_{k=0}^{K} \tilde{\beta}_k u^A(\phi(\hat{j}, k), j, k).$$

Whenever we refer to utility evaluated at prior belief about the principal, we suppress the argument $\tilde{\beta}$. The principal's interim utility with respect to the prior belief about the agent, given report $\hat{k}$, takes the form:

$$U^P_\phi(\hat{k}, k) := \int_0^J u^P_\phi(\phi(j, \hat{k}), j, k) dF(j).$$

We say an allocation ϕ is $\tilde{\beta}$-feasible if and only if it is $\tilde{\beta}$-incentive-compatible and $\tilde{\beta}$-individually rational; that is, it holds that $\forall k \in supp(\tilde{\beta})$ and $\forall j \in \Theta^A$:

$$(PIC)_k \qquad k \in \arg\max_{\hat{k} \in \Theta^P} U^P_\phi(\hat{k}, k),$$

$$(AIC)^j_{\tilde{\beta}} \qquad j \in \arg\max_{\hat{j} \in \Theta^A} U^A_\phi(\hat{j}, j, \tilde{\beta}),$$

$$(AIR)^j_{\tilde{\beta}} \qquad U^A_\phi(j, j, \tilde{\beta}) \geq U^A_D(j, \tilde{\beta}).$$

Whenever it is clear from the context that we consider the concept of β-incentive compatibility and β-individual rationality, we use the terms incentive compatibility and individual rationality. Note, $\tilde{\beta}$ might be a point belief about a certain principal type. If ϕ is $\tilde{\beta}$-feasible, where $\tilde{\beta}$ is a point belief about principal type k, we say that ϕ is $(\tilde{\beta}_k = 1)$-feasible. We say that an allocation is safe if and only if it is feasible for any possible point belief about the principal's type. That is, $\forall (j, k) \in \Theta$:

$$(PIC)_k \qquad k \in \arg\max_{\hat{k} \in \Theta^P} U^P_\phi(\hat{k}, k),$$

$$(AIC)^j_k \qquad j \in \arg\max_{\hat{j} \in \Theta^A} U^A_\phi(\hat{j}, j, k),$$

$$(AIR)^j_k \qquad U^A_\phi(j, j, k) \geq U^A_D(j, k).$$

Whenever it is clear from the context that incentive compatibility is satisfied, we suppress the argument $\hat{k}$ and $\hat{j}$ in the utility functions. Moreover, whenever it follows from the context that a certain constraint is imposed for all types of a given player, we suppress the script in the definition. For example, the constraints $(AIC)^j_{\tilde{\beta}}\ \ \forall j \in \Theta^A$ are denoted by $(AIC)_{\tilde{\beta}}$.

As a special instance, we denote the allocation given principal type k, as principal type k's menu:

$$\phi(\cdot, k) = (q_{\phi(\cdot,k)}(\cdot, k), q^{A,P}_{\phi(\cdot,k)}(\cdot, k), t_{\phi(\cdot,k)}(\cdot, k)) : \Theta^A \to [0,1]^2 \times [0,1] \times \mathbb{R}^2$$

The side mechanism that corresponds to $\phi(\cdot, k)$ specifies no action for the principal and is denoted as principal type k's menu (offer). If it is clear from the context that the we focus on a menu, we denote principal type k's outcomes by $t^P_{\phi(\cdot,k)}$ and $q^P_{\phi(\cdot,k)}$.

Because the collusive play of the auction can payoff-equivalently replicate the non-cooperative play, it is without loss of generality to focus attention on those $\tilde{\beta}$-feasible allocations which do not induce the non-cooperative play of the auction: Take any $\tilde{\beta}$-feasible allocation ϕ' such that for at least one type profile the SPA is played non-cooperatively, i.e., $q^{A,P}_{\phi'}(\cdot) > 0$. Then there exists some $\tilde{\beta}$-feasible allocation that does not induce the non-cooperative play of the SPA and implies the same utility for all types of both players.

Lemma 2.1. *It is without loss of generality to focus attention on allocations that are such that the SPA is never played non-cooperatively.*

A perfect Bayesian equilibrium for the grand game specifies (i) for each type of the principal an optimal mechanism proposal, (ii) for each side mechanism a belief $\tilde{\beta}$ about the principal's type, which is computed by Bayes' rule whenever possible, and (iii) for the continuation game starting with the proposal of any $m \in M$, a $\tilde{\beta}$-feasible allocation implemented by m.

The principle of inscrutability (see Myerson, 1983) applies: By the revelation principle, any allocation ϕ being induced by the perfect Bayesian equilibrium play of the grand game is β-incentive-compatible. Because the agent's payoff from the non-cooperative play of the SPA is linear in the belief about the principal, ϕ is β-feasible. Therefore ϕ can be induced through a pooling equilibrium, in which every principal type proposes a mechanism directly implementing ϕ.

A β-feasible allocation ϕ thus can be supported in a perfect Bayesian equilibrium if and only if for any $m' \in M$ there exists a belief $\tilde{\beta}$, such that m' induces a $\tilde{\beta}$-feasible allocation, ϕ', and $U^P_\phi(k) \geq U^P_{\phi'}(k)$ for all $k \in \Theta^P$.

Analysis

The considered three-stage collusive interaction falls into the class of informed-principal problems under common values. The common values feature emerges because the agent's outside option depends on the principal's type. This dependency makes the problem interesting. Different types of the principal prefer different allocations, and via the outside option, this information is payoff relevant for the agent. These two main features of our model are best illustrated by considering the principal's common-knowledge payoff. Type k's common-knowledge payoff is induced by his most preferred $\tilde{\beta}_k = 1$-feasible menu $\phi^c(\cdot, k)$.

Definition 2.1. *The common-knowledge allocation, ϕ^c, is the solution to the following class of maximization problems. For each $k \in \Theta^P$, $\phi^c(\cdot, k)$ is defined as solution to the problem*

$$\max_{\phi(\cdot,k)} U^P_\phi(k) \qquad (P^c)_k$$

such that $\forall j \in \Theta^A$ *:*

$$(AIR)_k^j, (AIC)_k^j, (BB)_{j,k}.$$

That is, $\phi^c(\cdot, k)$ induces for principal type k the largest feasible payoff in the hypothetical situation in which his type was commonly known.

Definition 2.2. *We say that allocation ϕ is efficient if, for any realized type profile, the bidder with the higher valuation receives the right. Else the allocation is inefficient.*

Remark 2.1. *ϕ^c is an efficient allocation, that is, $\forall (j,k) \in \Theta$:*

$$q^P_{\phi^c}(j,k) = \mathbb{1}[\theta_k \geq j], \qquad q^A_{\phi^c}(j,k) = 1 - q^P_{\phi^c}(j,k),$$
$$t^P_{\phi^c}(j,k) = -\mathbb{1}[\theta_k < j]\theta_k, \qquad t^A_{\phi^c}(j,k) = -t^P_{\phi^c}(j,k).$$

According to ϕ^c and given type k, the agent receives the right at transfer payment θ_k if and only if his valuation is above θ_k. Otherwise both the agent and the principal pay transfer zero and the principal receives the right. By this, principal type k can extract the entire collusive surplus - i.e., the bidders' payoff difference between playing the auction collusively and non-cooperatively. That is, $(AIR)_k^j$ is satisfied with equality for any $j \in \Theta^A$. Because $\phi^c(\cdot, k)$ is an efficient menu, the surplus is also maximized.

Because different types of the principal prefer different allocations, $q^P_{\phi^c(\cdot,k)} < q^P_{\phi^c(\cdot,k+1)}$. Moreover, ϕ^c does not satisfy the omitted (PIC) constraint. Precisely because of the common values feature of the considered collusive interaction, the agent's individual rationality constraints are eased in the type of the principal. As a consequence, a given principal type k prefers the allocation $\phi^c(\cdot, k+1)$ to the allocation $\phi^c(\cdot, k)$. Therefore, ϕ^c is not a β-feasible allocation[20] and cannot be implemented in equilibrium.

A safe allocation is feasible for any belief about the principal. The principal's best safe allocation is termed RSW (Rothschild-Stigliz-Wilson) allocation, introduced by Maskin and Tirole, 1992. By offering a mechanism which implements the best safe allocation from his perspective, a given principal type, say $\check{k}$, can secure himself a minimal payoff.

Definition 2.3. *The RSW allocation, ϕ^{RSW}, is the solution to the following class of maximization problems. For each $\check{k} \in \Theta^P$ we define the problem:*

$$\max_{\phi} U^P_{\phi}(\check{k}) \qquad (P1)_{\check{k}}$$

such that $\forall j, k \in \Theta$ *:*

$$(AIR)_k^j, (AIC)_k^j, (PIC)_k, (BB)_{j,k}.$$

That is, each principal type maximizes his utility subject to the constraint that the resulting allocation is $(\tilde{\beta}_k = 1)$-feasible for all $k \in \Theta^P$. ϕ^{RSW} is the solution to any of these problems. Denote the objects being induced by ϕ^{RSW} as $\{U^i_{RSW}(\cdot), q^i_{RSW}(\cdot), t^i_{RSW}(\cdot)\}_{i\in\{A,P\}}$.

The RSW allocation is inefficient in favor of the principal.

[20]Pathological cases in which ϕ^c is β-feasible are ruled out by assumption 2.

Lemma 2.2. *Whenever $F(\theta_k) < 1$ and $k > 0$, then $q^P_{RSW}(k) > F(\theta_k)$.*

If there are at least two principal types that receive the right with certainty, we say that there exists a pool at the top.

Definition 2.4. *Consider the set $\{k \in \Theta^P | q^P_{RSW}(k) = 1\}$. If it includes at least two types, we call it the pool at the top. We denote its lowest element by $\underline{K}$.*

In the appendix (section A.2), we provide a detailed derivation of the RSW allocation. Lemma 2.2 implies that the RSW allocation is inefficient in favor of the principal. The RSW allocation satisfies the same constraints as ϕ^c, and in addition, (PIC) (i.e., any type k prefers $\phi^{RSW}(\cdot, k)$ to $\phi^{RSW}(\cdot, k+1)$). Thus, principal types separate. Types differ in their valuation for the good. Hence, the screening parameter is $q^P_{RSW}(\cdot)$. Because ϕ^c is efficient and does not satisfy (PIC), ϕ^{RSW} is such that principal type k receives the right with probability larger than $F(\theta_k)$. Thereby, some agent types with valuation above θ_k do not receive the right, given type k. That is, separation introduces inefficiencies.

By construction, ϕ^{RSW} is $\tilde{\beta}$-feasible for any $\tilde{\beta}$. Therefore, a β-feasible allocation can only be supported in equilibrium if every principal type receives weakly larger payoff than $U^P_{RSW}(\cdot)$. Otherwise at least one type deviates by proposing a mechanism which implements the RSW allocation.

To argue that the converse is also true - i.e., every β-feasible allocation that induces every type of the principal weakly larger payoff than ϕ^{RSW} is an equilibrium allocation - we build upon a theorem of Mylovanov and Tröger (2015).

Lemma 2.3. *The methodology of Mylovanov and Tröger (2015) applies to our setting.*

Mylovanov and Tröger (2015) define a secured payoff for each type of the principal. In our setup, the secured payoff is implemented by the RSW allocation.

Lemma 2.4. *In the SPA, the secured payoff defined in Mylovanov and Tröger (2015) is implemented by the RSW allocation.*

If the vector consisting of each type's secured payoff, the secured-payoff vector, is undominated (from the principal's perspective) for some belief about the principal, then an allocation can be supported in equilibrium if and only if it is β-feasible and leaves every type of the principal with weakly larger utility than his secured payoff.

Definition 2.5. *Fix a belief $\tilde{\beta}$. An allocation ϕ is $\tilde{\beta}$-undominated if there exist positive welfare weights $\{z_k\}_{k=0}^K$, with $\sum_{k=0}^K z_k = 1$, $z_k > 0$ if $k \notin supp(\tilde{\beta})$ and $\sum\limits_{k \in supp(\tilde{\beta})} z_k > 0$,[21] such that ϕ solves:*

$$\max_{\phi} \sum_{k=0}^{K} z_k U^P_{\phi}(k) \tag{P2}$$

subject to:

$$\phi \text{ is } \tilde{\beta}\text{-feasible}$$

[21] Because we are in a quasilinear utility environment, welfare weights may equal zero for a type in the support of the belief.

We prove that there exists a belief, β^*, such that the RSW allocation is undominated. Note, this belief is not required to have full support, and in our setting β^* does not in general have full support.[22]

Lemma 2.5. *There exists β^* such that the RSW allocation is undominated.*

For fixed belief $\tilde{\beta}$ and fixed welfare weights, problems like $(P2)$ can be solved by carefully relaxing the constraints and rewriting the problem. Eventually the objective consists of (generalized) virtual valuations. The allotment policy of the optimal allocation can be identified as the point-wise maximizer of the objective (e.g., see Ledyard and Palfrey, 2007). We take the reverse approach. For a given claimed optimal allotment policy, $q_{RSW}(\cdot)$, we construct a belief β^* and welfare weights such that the resulting virtual valuations imply that $q_{RSW}(\cdot)$ is indeed a point-wise maximizer of the (constructed) objective, and the associated allocation, ϕ^{RSW}, is the optimal solution of the original problem $(P2)$, given $\tilde{\beta} = \beta^*$ and some welfare weights.

Given lemma 2.5, the next proposition follows by applying Mylovanov and Tröger, 2015.

Proposition 2.1. *An allocation can be supported in equilibrium if and only if it is β-feasible and leaves every type of the principal with weakly larger payoff than $U^P_{RSW}(\cdot)$.*

A direct consequence of proposition 2.1 is that the RSW allocation is always an equilibrium allocation.

We are interested in whether efficient collusion is included in the large class of equilibria allocations. The concept of β-undominated (equilibrium) allocations turns out to be useful for our analysis. Because $supp(\beta) = \Theta^P$, the definition 2.5 becomes:

Definition 2.6. *An allocation ϕ is β-undominated, or undominated, if there exist weakly positive welfare weights $\{z_k\}_{k=0}^K$ (with $z_k \geq 0\ \forall k \in \Theta^P$ and $\sum_{k=0}^K z_k = 1$) such that ϕ solves*

$$\max_{\phi} \sum_{k=0}^{K} z_k U^P_{\phi}(k) \qquad (\tilde{P}3)_z$$

subject to:

$$\phi \text{ is } \beta\text{-feasible}$$

Whenever ϕ is a solution of $(P3)_z$, which is defined similarly to problem $(\tilde{P}3)_z$ but augmented by the constraint

$$(PIR)_k \quad U^P_{\phi}(k) \geq U^P_{RSW}(k) \quad \forall k \in \Theta^P$$

we say that ϕ is an undominated equilibrium allocation.

[22]Given their setting, Maskin and Tirole (1992) prove (theorem 1^*) that whenever the RSW allocation is undominated (from the principal's perspective) for some belief with full support, any β-feasible allocation which leaves every principal type with weakly larger payoff than the RSW allocation can be supported in equilibrium. Moreover, they state conditions under which the RSW allocation is indeed undominated for full support beliefs (proposition 14). Not all of their conditions are satisfied in our setup. Most importantly, their sorting assumption does not in general apply in our setting. Consequently, we cannot directly apply their methodology. Indeed, whenever the sorting assumption fails in our setting (which happens if there is a pool at the top), the RSW allocation is dominated for all beliefs with full support.

Problems $(\tilde{P}3)_z$ and $(P3)_z$ are instances of characterizing interim incentive-efficient allocations in a quasilinear, independent-private-values environment. Because the bidders' outside options are convex in their types - (PIR) can be treated as the principal's individual rationality constraint - individual rationality constraints might bind for multiple types of a given player.[23]

An important benchmark is ex-ante optimal allocations. An ex-ante optimal allocation would result if the principal were able to propose the side mechanism before learning his type. For given prior β, he solves the problem of maximizing his expected utility among all allocations that are β-feasible.

Definition 2.7. *An ex-ante optimal allocation is a β-undominated allocation with $z_k = \beta_k$ for all $k \in \Theta^P$. That is, it is a solution to $(\tilde{P}3)_z$.*

An ex-ante optimal allocation corresponds to the hypothetical situation in which an uninformed third party proposes the side mechanism in order to maximize bidder P's (expected) payoff.

Proposition 2.2. *Every ex-ante optimal allocation implies efficient collusion.*

The intuition behind this result is as follows: The non-cooperative play of the SPA allots the good efficiently. If collusion is efficient, the collusive play leads to the same allotment of the good. By incentive compatibility, the difference between the agent's collusive payoff and non-cooperative payoff is necessarily a type-independent constant. By setting this constant to zero, the principal can extract the entire collusive surplus. Because collusion is efficient, the surplus is also maximized.

The form of the resulting ex-ante optimal allocation, ϕ^e, is known to the literature. Mailath and Zemsky, 1991 consider the case of a third party organizing collusion in a SPA. A characterization of the class of efficient allocations for any division of the collusive surplus between the cartel members is given. For the boundary case in which one member receives all surplus, ϕ^e results. However, Mailath and Zemsky, 1991 analyze a different problem and therefore do not show that ϕ^e is the ex-ante optimal allocation of this certain cartel member.[24]

Although ϕ^e allows the principal to absorb the entire collusive surplus, it cannot always be supported in equilibrium. Fix any undominated allocation ϕ. By standard mechanism design arguments (see, e.g, Ledyard and Palfrey, 2007), it follows that the principal's payoff can be represented in a compact form. The next lemma states this form for the case that the principal's upward adjacent incentive constraints hold with equality.[25] The most preferred allocation of the largest principal type falls into this class. As can be seen in the subsequent analysis, this is the most relevant undominated allocation for the purpose of this paper.

[23] To characterize the set of allocations to the degree necessary for the purpose of our paper, we employ some ideas of Ledyard and Palfrey, 2007. However, in contrast to the latter paper, we allow for cases in which the bidders' individual rationality constraints bind for more than one type of a given player.

[24] As minor technical (but in our opinion, interesting) detail: When proving that ϕ^e is a solution to $(\tilde{P}3)_z$, every agent type $(AIR)^j$ constraint is binding.

[25] If the upward adjacent incentive constraints do not hold with equality, the payoffs look similar as in lemma 2.6, but the form of $v_\phi^P(k)$ is cumbersome. For readability, we omit these cases in the statement of the lemma.

Lemma 2.6. *Any undominated allocation ϕ, according to which the principal's upward adjacent incentive constraints hold with equality, induces the following payoffs:*

$$U_\phi^P(k) = v_\phi^P(k) + U_\phi^P(0) \quad \forall k \in \Theta^P,$$

with

$$v_\phi^P(k) := U_\phi^P(k) - U_\phi^P(0) = \sum_{v=1}^{k} q_\phi^P(v)(\theta_v - \theta_{v-1}) \quad \forall k \in \Theta^P,$$

$$v_\phi^A(j) := U_\phi^A(j) - U_\phi^A(0) = \int_0^j q_\phi^A(v)dv \quad \forall j \in \Theta^A,$$

and

$$U_\phi^P(0) = \mathbb{E}_j[q_\phi^A(j)j - v_\phi^A(j)] + \mathbb{E}_k[q_\phi^P(k)\theta_k - v_\phi^P(k)] - \max_{j\in\Theta^A}[U_D^A(j) - v_\phi^A(j)].$$

Given the above necessary condition, we provide in section A.9 sufficient conditions assuring that an allocation is indeed undominated. We term $v_\phi^P(k)$ the information rent of principal type k and $U_\phi^P(0)$ the absorbed collusive surplus.

The principal is a player of the game he proposes. Therefore, any β-incentive-compatible allocation yields him an information rent. Because the screening parameter is the probability of receiving the right, $v_\phi^P(k)$ is larger the larger $q_\phi^P(v)$ is for any $v \leq k$. In this sense, the principal prefers inefficient allocations in favor of him. In addition, the principal designs the mechanism. By this he can extract part of the collusive surplus and receives a constant payment, $U_\phi^P(0)$. If collusion is efficient, this surplus is maximized and fully extracted. If collusion is inefficient the surplus is not maximized. Moreover, the agent receives a collusive rent equal to $\max_{j\in\Theta^A}[U_D^A(j) - v_\phi^A(j)]$, which assures his participation.

The RSW allocation is inefficient in favor of the principal, given any type of the principal. If collusion is efficient, then $q_{\phi^e}^P(k) \leq q_{\phi^{RSW}}^P(k)$, and thus $U_{RSW}^P(k) - U_{\phi^e}^P(k)$ is increasing in k. Therefore, efficient collusion can be supported in equilibrium if and only if the highest type of the principal K prefers ϕ^e to ϕ^{RSW}.

Lemma 2.7. *Efficient collusion can be supported in equilibrium if and only if $U_{\phi^e}^P(K) \geq U_{RSW}^P(K)$.*

We first verify that if types are close to being symmetric, even the highest type prefers efficient collusion to any other β-feasible allocation.

Definition 2.8. *Suppose $\theta_K + \Delta = J$, with $\Delta := \theta_k - \theta_{k-1} = \frac{\theta_K}{\mathrm{K}}$. Suppose that $f(\cdot)$ is differentiable and $\beta_k = F(\Delta + \theta_k) - F(\theta_k)$ for any $k > 0$. For any $k > 0$, define $\tilde{f}(k) := F(\Delta + \theta_k) - F(\theta_k) - \Delta f(\theta_k)$.*
If $\tilde{f}(k)$ is non-negative, $\max_{k>0} \tilde{f}(k) \leq \delta$ and such that $\frac{\tilde{f}(k)}{f(k)}$ is non-decreasing in k, we say that $F(\cdot)$ and $B(\cdot)$ are δ-close distributions.

We show that ϕ^e is a solution to problem $(\tilde{P}3)_z$ with $z_K = 1$ whenever $F(\cdot)$ and $B(\cdot)$ are δ-close distributions, with δ and/or Δ being sufficiently close to zero. That is, the principal's type distribution is a sufficiently good discrete approximation of the agent's type distribution; the remaining conditions in definition 2.8 are of technical nature.[26]

Lemma 2.8. *Assume that $F(\cdot)$ and $B(\cdot)$ are δ-close distributions. If δ is sufficiently close to zero, then ϕ^e is an equilibrium allocation.*

Suppose the highest principal type, K, were able to decide upon the collusive allocation ϕ. That is, he maximizes $v_\phi^P(K) + U_\phi^P(0) - \max_{j\in\Theta^A}[U_D^A(j) - v_\phi^A(j)]$. Let him start at an efficient allocation ϕ^e. Suppose he introduces inefficiencies on the allotment policy in favor of the principal such that his information rent increases. In turn, the information rent of the highest agent type, J, immediately decreases to a value below the information rent the latter receives from the non-cooperative play of the SPA. To assure that the agent is still willing to participate in the collusive agreement, the latter, independently of his type, receives a lump-sum payment equal to the loss of J's information rent. Hence, type K optimally maximizes the sum of his and J's information rent. Since both bidders are close to being symmetric, a symmetric allotment policy is desired.

Having established this feature, we interpret $U_{\phi^e}^P(0)$ as the seller's forgone revenue being extracted by the principal. This revenue increases, the stronger bidder P is in stochastic terms. Because $U_{RSW}^P(K)$ is independent of β, the difference between $U_{\phi^e}^P(K)$ and $U_{RSW}^P(K)$ increases if bidder P becomes stronger. Hence, if ϕ^e can be supported in equilibrium for some β with associated distribution $B(\cdot)$, it can also be supported for any $\tilde{\beta}$ whose associated distribution, $\tilde{B}(\cdot)$, first-order stochastically dominates $B(\cdot)$.

However, efficient collusion cannot always be supported in equilibrium. If bidder P is much weaker than bidder A, and the former has large probability mass on low types, then the collusive surplus is rather low. In such an instance, from type K's perspective, maximizing and extracting the collusive surplus is less desirable than maximizing the information rent. As a consequence, K prefers inefficient allocations, such as ϕ^{RSW}, to ϕ^e.

Proposition 2.3. *For any given distribution of the agent's type, there exists a distribution of the principal's type, $\hat{B}(\cdot)$,[27] such that the highest principal type is indifferent between $U_{RSW}^P(K)$ and $U_{\phi^e}^P(K)$.*

Moreover, principal type K's payoff induced by efficient collusion increases if the distribution of the principal's type increases in the sense of first-order stochastic domination. Hence, if $B(\cdot)$ first-order stochastically dominates $\hat{B}(\cdot)$, then efficient collusion can arise in equilibrium. If $B(\cdot)$ is first order stochastically-dominated by $\hat{B}(\cdot)$, then efficient collusion cannot be supported in equilibrium.

An immediate consequence of lemma 2.8 and proposition 2.3 is that if both bidders can be ordered according to first-order stochastic dominance and the stronger of

[26]If the technical conditions are violated, then we would suspect that efficient collusion still can be supported as equilibrium of the game. Let ϕ_{z_K} denote the solution to problem $(\tilde{P3})_z$ with $z_K = 1$. For any ϕ, $U_\phi^P(K)$ is continuous in $B(\cdot)$ and $F(\cdot)$. Thus, though $\phi_{z_K} \neq \phi^e$, if $B(\cdot)$ is sufficiently close to $F(\cdot)$ then $\max_{k\in\{1,\ldots,K\}} |U_{\phi^e}^P(K) - U_{\phi_{z_K}}^P(K)|$ is sufficiently small. Hence, the principal's utility induced by ϕ^e also dominates the RSW payoff, implying ϕ^e to be an equilibrium allocation.

[27]$\hat{B}(\cdot)$ is not unique. In the proof of lemma 2.3 we construct $\hat{B}(\cdot)$ as a convex combination between a distribution that has sufficiently large probability mass on type 0, and arbitrary little on all other types, say $B_0(\cdot)$, and one that has sufficiently large probability mass on type K. However, one also can construct $\hat{B}(\cdot)$ by taking a convex combination between B_0 and a distribution that is close to the agent's type distribution, in the sense of definition 2.8.

them is the principal,[28] efficient collusion can be supported in equilibrium. That is, if both bidders are not too asymmetric or can be ordered according to first-order stochastic dominance, then one can always find a distribution of bargaining power such that the cartel can achieve its joint profit-maximizing benchmark of efficient collusion.

These are our main results concerning collusion in the SPA. In section 2.5.1 we apply the Intuitive Criterion to refine the set of equilibria.

2.4 Collusion in the FPA

We stick to the notation introduced above. In particular, the set of the principal's side mechanisms M is the same as in the case of collusion in the SPA. What changes are the payoffs induced by the non-cooperative play of the auction, $U_D^P(\cdot)$ and $U_D^A(\cdot)$.[29] The non-cooperative equilibrium play of the FPA depends on the bidders' beliefs at the beginning of stage $t = 3$, and so do $U_D^P(\cdot)$ and $U_D^A(\cdot)$. Hence, in the FPA, the bidders' payoffs induced by any (collusive) allocation ϕ that results from their interaction differs from the case of collusion in the SPA. However, this difference exclusively affects the non-cooperative outcome $q_\phi^{A,P}(\cdot)$. Without loss of generality, we assume that in case the non-cooperative outcome results, this event is commonly known by both bidders.[30]

Preliminary Observations Let us start our analysis by defining a feasible allocation in the FPA.

Definition 2.9. *Suppose collusion is in the FPA and the agent holds belief $\tilde{\beta}$ about the principal. We say that an allocation ϕ is $\tilde{\beta}$-feasible if there exists an off-path belief about the agent's type such that the allocation can be implemented by a mechanism that is accepted by all types of the agent. We say that this certain off-path belief about the agent enforces ϕ.*

Continuation Game after period 1 Consider the continuation game starting with the beginning of period $t = 2$ where the agent decides whether to accept or to reject m, the proposed collusive side mechanism. The agent's belief about the principal, $\tilde{\beta}$, might differ from the prior. If a continuation equilibrium exists (more on this below), the revelation principle implies that the bidders' interaction results in a (collusive) allocation, ϕ, which is $\tilde{\beta}$-incentive-compatible. Moreover, ϕ can be implemented by a mechanism, $\hat{m}$, that is accepted by every agent type. Either, the considered continuation equilibrium play that induces ϕ is such that every agent type accepts m. In this case, $\hat{m}$ can be chosen equals m, and ϕ is enforced by the off-path belief that would result if m were rejected. Or, the continuation equilibrium play that induces ϕ is such that m is rejected by some agent types, that is, $q_\phi^{A,P}(\cdot) > 0$ for some type profiles. In this case, $\hat{m}$ can, for example, be chosen as

[28] Of course, the principal's type space is required to be a discrete approximation of a continuous type space as in definition 2.8.

[29] We slightly abuse notation by redefining in this section $U_D^A(\cdot)$ as the agent's payoff induced by the non-cooperative play of the FPA.

[30] This assumption is equivalent to not enlarging M by those mechanisms that allow for recommended actions in the play of the FPA. Because of the side transfers, the bidders' types can be directly screened without relying on the play of some communication equilibrium of the FPA. Indeed, the arguments presented in the subsequent analysis do survive an enlargement, in the above sense, of M.

a revelation mechanism that directly implements ϕ. The commonly known belief the principal holds about the agent after observing the outcome $q_{\phi}^{A,P}(\cdot)$, is equal to the on-path belief the former holds about the latter after observing the rejection of m. Moreover, because ϕ is $\tilde{\beta}$-incentive-compatible, the considered belief about the agent necessarily enforces ϕ. As a consequence, any allocation that is induced through some (continuation) equilibrium play is $\tilde{\beta}$-feasible.

Non-cooperative play of the FPA We focus attention on equilibria of the FPA that feature non-dominated bidding. That is, no type of a bidder submits a bid above his type. Given this, we show that any $\tilde{\beta}$-feasible allocation can be enforced by a degenerate off-path belief concentrated on the highest agent type.

Suppose the agent holds arbitrary belief $\tilde{\beta}$ about the principal and rejects the proposed mechanism m that implements a $\tilde{\beta}$-feasible allocation. Assume this behavior results in a degenerate (off-path) belief concentrated on the highest agent type.

Lemma 2.9. *The non-cooperative play of the FPA, given belief $\tilde{\beta}$ about the principal and a degenerate belief concentrated on the highest agent type J,*[31] *induces agent type j the payoff*

$$U_D^A(j,\tilde{\beta}) = \max_{\check{k}\in\Theta^P} \tilde{B}(\check{k})(j-\theta_{\check{k}}).$$

By lemma 2.9, the play of the FPA is from the agent's point of view as if every type of the principal submits a bid equal to his type. That is, principal types' bids are maximized. Hence, a degenerate belief concentrated on agent type J is the worst off-path belief from every agent types' perspective. Thus, any $\tilde{\beta}$-feasible allocation can be enforced by a degenerate off-path belief concentrated on agent type J.

A perfect Bayesian equilibrium for the grand game specifies (i) an optimal mechanism proposal for each type of the principal, (ii) for each mechanism m a belief $\tilde{\beta}$ about the principal type, which is computed by Bayes' rule whenever possible; and (iii) for the continuation game starting with the proposal of m a $\tilde{\beta}$-feasible allocation implemented by m.

Note that Myerson (1983)'s principle of inscrutability applies to the FPA. Any $\tilde{\beta}$-feasible allocation can be enforced by a degenerate off-path belief concentrated on agent type J. By the revelation principle, any allocation ϕ being induced by the perfect Bayesian equilibrium play of the grand game is β-incentive-compatible. Because the agent's payoff from the non-cooperative play of the FPA with degenerate belief concentrated on agent type J is convex in the belief about the principal, ϕ is β-feasible.[32] Therefore ϕ can be induced through a pooling equilibrium in which every principal type proposes a mechanism directly implementing ϕ.

[31] In case there are multiple continuation equilibria, we choose the following: Every principal type weakly below $\arg\max_{\check{k}\in\Theta^P} \tilde{B}(\check{k})(J-\theta_{\check{k}})$ submits a bid equal to his type. This approach assures that any $\tilde{\beta}$-feasible allocation can be enforced by an off-path belief about agent type J. We show in the subsequent analysis that the chosen continuation equilibrium is part of the unique perfect Bayesian equilibrium of the grand game.

[32] This means, concerning the non-cooperative play of the FPA, the agent can only be harmed from being less informed about the principal's type. Or, let M^e be the set of those game forms that occur with positive probability in some hypothetical separating equilibrium. Denote the probability that m^e with $m^e \in M^e$ occurs in equilibrium by $Pr(m^e) > 0$ and by β_{m^e} the resulting posterior about the principal. Let $B_{m^e}(\cdot)$ be the associated distribution. Then, $\sum_{m^e\in M^e} Pr(m^e) \max_{\check{k}\in\Theta^P} B_{m^e}(\check{k})(j-\theta_{\check{k}}) \geq \max_{\check{k}\in\Theta^P} B(\hat{k})(j-\theta_{\check{k}})$. That is, the channel identified by Dequiedt, 2006 and Celik and Peters, 2011 does not apply.

A β-feasible allocation ϕ thus can be supported in a perfect Bayesian equilibrium if and only if for any $m' \in M$ there exist $\tilde{\beta}$ such that m' induces a $\tilde{\beta}$-feasible allocation, ϕ', and $U^P_\phi(k) \geq U^P_{\phi'}(k)$ for all $k \in \Theta^P$.

Outline of the approach To apply the informed-principal methodology used to analyze collusion in the SPA, one must verify the existence of equilibrium of the non-cooperative play of the FPA for arbitrary beliefs, both about the principal and the agent.[33] To avoid this, we take a different approach to solve the considered three-stage collusive interaction. For the special cases in which $\tilde{\beta}$ is a point belief or the belief about the agent is a point belief, the non-cooperative play of the FPA can be analyzed without complications. We employ this feature and directly characterize an equilibrium of the grand game. Afterwards, we verify that it is unique (in terms of payoffs). To be more precise, we verify that the following play can be supported in equilibrium: Every type of the principal proposes the same mechanism, $\tilde{m}$, and the agent accepts the proposal. If the agent deviates and rejects the mechanism, the FPA is played non-cooperatively with degenerate off-path belief concentrated on the highest agent type. If the principal deviates and proposes m different from $\tilde{m}$, the continuation game is played with the agent holding a degenerate off-path belief concentrated on the lowest principal type. That is, any principal's deviation induces some $(\tilde{\beta}_0 = 1)$-feasible allocation.

Throughout the subsequent analysis, we remark relations to the informed-principal methodology in footnotes.

What makes the analyzed problem interesting is the interplay between (i) the principal's large set of side mechanisms and (ii) the dependence of the agent's bid in the non-cooperative play of the FPA on $\tilde{\beta}$.

Let us abuse notation a bit and indicate by $\phi^{\tilde{\beta}}_m$ the allocation that is implemented by the side mechanism m with the agent holding arbitrary belief $\tilde{\beta}$ about the principal. This abuse of notation is justified, because, as a consequence of point (ii) above, side mechanisms that implement different allocations for different beliefs about the principal are relevant. One such side mechanism is $\tilde{m}$. It plays a prominent role in our analysis.

Mechanism $\tilde{m}$. *Denote the agent's action or report by $d \in \Theta^P$. Let the principal's action or report be $\hat{k} \in \Theta^P$. Given the report profile $(d, \hat{k}) = (\check{k}, k)$, $\tilde{m}$ specifies the collusive outcome $(q^i_{\tilde{m}}(\cdot), t^i_{\tilde{m}}(\cdot))_{i\in\{A,P\}}$ according to:*

$$(q^A_{\tilde{m}}(d = \check{k}, k), t^A_{\tilde{m}}(d = \check{k}, k)) = \begin{cases} (0, 0) & \text{if } k > \check{k} \,, \\ (1, \theta_{\check{k}}) & \text{if } k \leq \check{k} \,. \end{cases}$$

$$(q^P_{\tilde{m}}(d = \check{k}, k), t^P_{\tilde{m}}(d = \check{k}, k)) = \begin{cases} (1, 0) & \text{if } k > \check{k} \,, \\ (0, -\theta_{\check{k}}) & \text{if } k \leq \check{k} \,. \end{cases}$$

[33] With the beginning of period 2 the agent decides whether to accept or to reject the mechanism proposed by the principal. At the start of this continuation game, the agent might hold some arbitrary belief, $\tilde{\beta}$, about the principal. If the agent rejects the mechanism, the FPA is played non-cooperatively, possibly with the principal holding belief about the agent which differs from the prior. Arguing for existence of the equilibrium of the non-cooperative play of the FPA in such a setting is a difficult task. Given the literature, e.g., Krishna, 2002, we strongly believe that it is impossible without imposing more structure on the FPA (e.g., allow only for discrete bid increments). However, this comes at notational and computational-costs.

Note that $\tilde{m}$ is a simultaneous-move mechanism but not a direct revelation mechanism.

Lemma 2.10. *For any $\tilde{\beta}$, the play of $\tilde{m}$ is such that principal type k reports $\hat{k} = k$, every agent type receives payoff equal to his outside option, $U_D^A(j, \tilde{\beta})$, and any principal type k is left with weakly larger payoff than θ_k.*

As a consequence, for any $\tilde{\beta}$ the mechanism $\tilde{m}$ directly implements a $\tilde{\beta}$-feasible allocation, $\phi_{\tilde{m}}^{\tilde{\beta}}$, with $U^P_{\phi_{\tilde{m}}^{\tilde{\beta}}}(k) \geq \theta_k$.

As for intuition, the non-cooperative play of the FPA depends on $\tilde{\beta}$. For example, if this belief has high mass on low principal types, then bidder A submits a fairly low bid and wins the auction only against low types of bidder P, but at a low price. Conversely, if $\tilde{\beta}$ has high mass on high bidder P types, bidder A submits a rather high bid and wins the auction against high types of bidder P, but at a large price.

That is, if $\check{k}(j, \tilde{\beta})$ stands for $\arg\max_{\check{k} \in \Theta^P} \tilde{B}(\check{k})(j - \theta_{\check{k}})$, then the non-cooperative play of the FPA is from the agent's point of view as if the following allocation were induced:

$$q_D^A(j,k) = \mathbb{1}[\check{k}(j,\tilde{\beta}) \geq k], \quad t_D^A(j,k) = \mathbb{1}[\check{k}(j,\tilde{\beta}) \geq k]\theta_{\check{k}(j,\tilde{\beta})}$$

The mechanism $\tilde{m}$ exploits this famous bid-shading property of the FPA. $\tilde{m}$ corresponds to the following hypothetical situation: The principal constructs a pool of K+1 menus, with a generic menu indexed by $\check{k}$. Afterwards, it's the agent's turn to propose a menu out of the pool. Finally, the principal determines the allocation by picking an element from the proposed menu.

Each menu in the constructed pool consists of a price at which the principal can buy the right and a compensation. If the principal picks the compensation, he commits to stay out of the auction and the agent receives the right. In menu $\check{k}$ the price is equal to zero whereas the compensation is equal to $\theta_{\check{k}}$.

If principal type k chooses from menu $\check{k}$, he picks the compensation if and only if $k \leq \check{k}$. Thus, for any $\tilde{\beta}$, selecting the optimal menu is from the agent's perspective exactly the same problem as deciding upon the optimal bid in the non-cooperative play of the FPA. Hence, every agent type accepts $\tilde{m}$ and receives payoff equal to his outside option.

As a direct consequence of lemma 2.10, in any equilibrium of the grand game in which the continuation game that starts after the rejection of $\tilde{m}$ is played with degenerate off-path belief concentrated on agent type J, no type of the principal receives lower payoff than θ_k. This is the payoff principal type k receives from proposing $\tilde{m}$, when $\tilde{m}$ is played with the worst (off-path) belief, $\tilde{\beta}_0 = 1$, from any principal's type perspective, *given* the mechanism $\tilde{m}$.[34]

Note, by proposing $\tilde{m}$, large principal types secure themselves a huge payoff. As for comparison, let us consider the principal's common-knowledge payoff in the FPA.

[34]In the SPA, the principal's best safe payoff is $U^P_{RSW}(\cdot)$. Here, the best safe payoff, or the RSW allocation, yields all but the lowest principal type a lower payoff than θ_k and thus cannot be supported in equilibrium. In fact, one can show that θ_k is the principal's secured payoff, defined by Mylovanov and Tröger, 2015. Thus, the secured payoff is different from the best safe payoff.

Definition 2.10. *In the FPA, the common-knowledge allocation, $\tilde{\phi}^c$, is the solution to the following class of maximization problems. For each $k \in \Theta^P$, $\tilde{\phi}^c(\cdot, k)$ is defined as solution to the problem*

$$\max_{\phi(\cdot,k)} U^P_\phi(k) \qquad (\tilde{P}^c)_k$$

such that $\forall j \in \Theta^A$:

$$(AIC)^j_k, (BB)_{j,k},$$

$$(AIR)^j_k \quad U^A_{\phi(\cdot,k)}(j,k) \geq U^A_D(j, \tilde{\beta}_k = 1) = \mathbb{1}[j \geq k](j - \theta_k)$$

The difference to definition 2.1, which defines the common-knowledge payoff in the SPA, is that the $(AIR)_{\tilde{\beta}_k=1}$ constraints are evaluated at a different default game. However, the play of the FPA with point belief about principal type k and degenerate off-path belief concentrated on agent type J is such that principal type k submits a bid equal to his type. Hence, the agent's $(AIR)_{\tilde{\beta}_k=1}$ constraints are the same as in definition 2.1. By this, the common-knowledge allocation takes the same form as in the SPA.

Remark 2.2. *$\tilde{\phi}^c = \phi^c$. Hence, $\tilde{\phi}^c$ is an efficient allocation, and $U^P_{\tilde{\phi}^c(\cdot,k)}(k) = \theta_k$ for all $k \in \Theta^P$.*

Hence, $\tilde{m}$ leaves especially large principal types with weakly larger payoff than their common-knowledge payoff. As a consequence, the proposal can be supported as equilibrium, which takes the following form:

Every type of the principal proposes $\tilde{m}$. If the principal proposes a different mechanism, a degenerate off-path belief concentrated on the lowest principal type results. If the agent rejects $\tilde{m}$, a degenerate off-path belief concentrated on the highest agent type results.

Proposition 2.4. *There exists an equilibrium such that every principal type proposes $\tilde{m}$ or a mechanism that implements $\phi^\beta_{\tilde{m}}$.*

Any deviation, say m', of the principal results in a degenerate off-path belief concentrated on the lowest principal type, $\tilde{\beta}_0 = 1$. Given this, the non-cooperative play of the FPA, for any belief about the agent, is such that the agent submits a bid equal to θ_0 and wins the auction with certainty against principal type 0. Consequently the considered (off-path) continuation game has an equilibrium and we know about its properties. This knowledge is sufficient to verify that any $(\tilde{\beta}_0 = 1)$-feasible allocation leaves every type of the principal with payoff no larger than θ_k. As for intuition, even if all agent types reject the side mechanism m' and every principal type k, with $k > 0$, wins the auction at price $\theta_0 = 0$, no type receives payoff larger than his common-knowledge payoff, θ_k.[35]

The equilibrium is unique:

Theorem 2.1. *In any perfect Bayesian equilibrium of the grand game, the allocation implemented by the play of $\tilde{m}$ is the unique β-feasible equilibrium allocation.*

To establish the theorem, we build upon the next two lemmatas: We first restrict our attention to those equilibria in which the continuation game that starts after the

[35] To relate to the informed-principal methodology: We essentially show that the secured-payoff vector is undominated for the belief $\tilde{\beta}_0 = 1$.

agent rejected $\tilde{m}$ is played with a degenerate off-path belief concentrated on agent type J. We show that, for any prior β there exists only one allocation in terms of payoffs, which yields the largest principal type payoff no lower than θ_K. This is the allocation that is implemented by the play of $\tilde{m}$ under β. Every other β-feasible allocation leaves the largest principal type with strictly lower payoff than θ_K and therefore cannot be supported in equilibrium.

Lemma 2.11. *In any perfect Bayesian equilibrium of the grand game, being such that the rejection of $\tilde{m}$ is followed by a degenerate off-path belief concentrated on agent type J, the allocation implemented by the play of $\tilde{m}$ is the unique β-feasible equilibrium allocation.*

Given this, the following observation establishes uniqueness: Suppose there exists some hypothetical equilibrium which induces an allocation different from $\phi^{\beta}_{\tilde{m}}$. We know from the proof of lemma 2.11 that principal type K receives strictly lower payoff than θ_K in this equilibrium. Suppose principal type K deviates from such a hypothetical equilibrium by proposing $\tilde{m}_\epsilon$. This mechanism is a slightly perturbed version of $\tilde{m}$, such that every agent type receives ϵ-larger payoff than $\max_{\check{k}\in\Theta^P}\tilde{B}(\check{k})(j-\theta_{\check{k}})$, with ϵ positive, but arbitrarily small. In any perfect Bayesian equilibrium of the grand game, the continuation equilibrium that begins after the principal's proposal of $\tilde{m}_\epsilon$ is such that every agent type accepts $\tilde{m}_\epsilon$: Suppose to the contrary that some agent types, holding arbitrary off-path belief $\tilde{\beta}$ on the principal, reject $\tilde{m}_\epsilon$. Denote by $\underline{j}$ the lowest such agent type and by $\Pi^A(j)$ the agent's expected payoff from playing the FPA non-cooperatively against the principal, with the agent holding arbitrary belief about the principal, and the principal holding belief about the agent that only has positive mass on those types for which $\Pi^A(j) \geq \max_{\check{k}\in\Theta^P}\tilde{B}(\check{k})(j-\theta_{\check{k}})+\epsilon$. We show that, in any such equilibrium of the FPA in non-dominated bidding strategies[36] type $\underline{j}$ receives payoff $\Pi^A(\underline{j}) = \max_{\check{k}\in\Theta^P}\tilde{B}(\check{k})(\underline{j}-\theta_{\check{k}})$. Thus, $\underline{j}$ strictly prefers to accept $\tilde{m}_\epsilon$, leaving him with ϵ larger payoff. Hence, the continuation game that starts after the proposal of $\tilde{m}_\epsilon$ features a unique perfect Bayesian equilibrium: Every agent type accepts $\tilde{m}_\epsilon$. Therefore, an allocation that induces the highest principal type payoff strictly lower than θ_K cannot be supported in a perfect Bayesian equilibrium. K deviates by proposing $\tilde{m}_\epsilon$ and receives payoff arbitrarily close to $\tilde{m}$.

Lemma 2.12. *In any equilibrium of the grand game, the rejection of $\tilde{m}$ is followed by a degenerate off-path belief concentrated on agent type J.*

By this, the unique equilibrium in the FPA is inefficient, making it impossible for the cartel to achieve its joint profit-maximizing benchmark of efficient collusion.

Note that this equilibrium is robust in the sense that it survives refinement concepts. Concerning a deviation of the agent, it satisfies the concept of ratifiability introduced by Cramton and Palfrey, 1995 and also the Intuitive Criterion. Concerning a deviation of the principal, the equilibrium necessarily survives the Intuitive Criterion, because every type of the principal receives weakly larger payoff than his common-knowledge payoff.

Remark 2.3. *$\phi^{\beta}_{\tilde{m}}$ is ratifiable. Moreover, the equilibrium satisfies the Intuitive Criterion.*

[36] If there does not exist an equilibrium of the FPA, the unique continuation equilibrium is the one which every agent type accepts $\tilde{m}_\epsilon$.

The allocation being implemented by $\tilde{m}$, $\phi^{\beta}_{\tilde{m}}$, leaves every principal type with at least his common-knowledge payoff. Compared to collusion in the SPA ,this is an extremely large payoff for high principal types. Low types receive from $\phi^{\beta}_{\tilde{m}}$ larger payoff than from the RSW allocation, the best safe allocation in the SPA. However, compared to less inefficient equilibria allocations, low types are better off in the SPA.

As an implication, in the FPA, the distribution of the principal's utilities is fairly unequal. As a consequence, $\phi^{\beta}_{\tilde{m}}$ gives rise to large principal information rents and thus to large inefficiencies of the allotment policy in favor of the principal.

Hence, the least inefficient equilibrium of the SPA is less inefficient than the unique equilibrium of the FPA.

Definition 2.11. *The least inefficient equilibrium allocation in the SPA is defined as solution to:*

$$\max_{\phi} \mathrm{E}_{j,k}[U^P_{\phi}(k) + U^A_{\phi}(j)]$$

subject to:

$$\phi \text{ is } \beta\text{−feasible,}$$

$$U^P_{\phi}(k) \geq U^P_{RSW}(k) \quad \forall k \in \Theta^P$$

We show that if the cartel can achieve its least inefficient equilibrium, then collusive profits are larger in the SPA than in the FPA.

Proposition 2.5. *The least inefficient equilibrium if collusion is in the SPA is less inefficient than the unique equilibrium in case collusion is in a FPA.*

Both the allotment policy resulting from collusion in the FPA and the RSW allotment policy, defined for the case of collusion in the SPA, can be described by a weakly increasing sequence of threshold agent types. Principal type k receives the right if and only if the agent's type is below the threshold associated with type k. Now consider the allotment policy that corresponds to the element-wise minimum of these two sequences. Because both sequences induce an inefficient allotment policy - i.e., the threshold agent type associated with principal type k is larger than θ_k - we know that the allotment policy described by the element-wise minimum sequence is less inefficient than any of the original policies. Suppose that at least one element of the latter sequence is equal to the FPA threshold. Then, we show that the policy corresponding to the element-wise minimum sequence gives rise to an allocation that every principal type prefers to the RSW allocation. Hence, the element-wise minimum sequence can be implemented in an equilibrium of the SPA. By this we establish that the least inefficient equilibrium is necessarily less inefficient than the unique FPA equilibrium.

2.5 Extensions for Collusion in the SPA

In the next sections we give a characterization allocations that survive the Intuitive Criterion.

2.5.1 The Intuitive Criterion

In this section we characterize the set of equilibria surviving Cho and Kreps, 1987's Intuitive Criterion (henceforth CK criterion). We show that efficient collusion is never contained in this set. Since the CK criterion is a rather mild refinement concept, this implies that efficient collusion is only robust to a minimal form of signaling. Let us start by introducing some additional notation.

CK criterion

Consider some given equilibrium allocation which induces $\{U^p(k), q^p(k), t^p(k)\}_{k=0}^K$. Consider an off-path collusive side mechanism m'. We define by $M(k, \tilde{\beta}, m')$ the set of continuation payoffs of principal type k, induced by the play of the continuation game starting with the beginning of period $t = 2$. The continuation game begins with the agent's decision whether to accept m' while holding some (arbitrary) belief $\tilde{\beta}$ about the principal. Whenever the mechanism is rejected, we still hold fix the weakly dominant strategy equilibrium of the SPA. Note that $M(k, \tilde{\beta}, m')$ is a set, since it might be the case that the play of the collusive side mechanism m' has multiple equilibria.

We indicate by $\tilde{\beta}^1 \in B_1(m') := \{b \in \mathbb{R}_+^K | \sum_{k=0}^K b_k = 1, supp(b) = \Xi_1(m')\}$ a belief about the principal that is restricted to have support only in the set Ξ_1, being defined by

$$\Xi_1(m') := \{k \in \Theta^p | \exists \tilde{\beta} \in B_0 : \max M(k, \tilde{\beta}, m') \geq U^p(k)\}.$$

We also define the following set:

$$\Xi_2(m') := \{k \in \Xi_1(m') | \min M(k, \tilde{\beta}^1, m') > U^p(k) \ \ \forall \tilde{\beta}^1 \in B_1(m')\}.$$

An equilibrium allocation fails to satisfy the CK criterion if and only if there exists an m' such that $\Xi_2(m')$ is nonempty.

We show that an equilibrium allocation satisfies the CK criterion if and only if there does not exist a separation menu for any principal type $k \in \Theta^p$.

Definition 2.12. *Fix principal types k and $k-1$ with equilibrium payoffs $U^p(k)$ and $U^p(k-1)$. A menu $\phi(\cdot, k)$ is called separation menu for type k if it satisfies:*

$$(AIR)_k^j, (AIC)_k^j, (BB)_{j,k} \ \ \forall j \in \Theta^A,$$
$$U^p_{\phi(\cdot,k)}(k-1) < U^p(k-1),$$
$$U^p_{\phi(\cdot,k)}(k) > U^p(k).$$

Denote the menu by ϕ_k^S and let $t^p_{\phi_k^S}, q^p_{\phi_k^S}$ stand for the principal's expected transfer payment and allotment probability, conditional that the menu is unanimously accepted by all agent types.

A separation menu for principal type k satisfies the agent's individual rationality and incentive constraints given a point belief about the principal type k. Moreover, in contrast to the principal type k, $k-1$ does not prefer ϕ_k^S to the equilibrium allocation.

The concept of a separation menu is useful to determine whether an equilibrium allocation satisfies the CK criterion. However, it should be noted that we do not prove that a separation menu is an off-path side mechanism making the equilibrium fail the CK criterion. We do prove, however, that the failure of the CK criterion implies the existence of a separating menu. Moreover, we show that the existence of a separation menu implies the existence of some side mechanism, which in turn implies the failure of the CK criterion.

We sketch the only if part of the proof as follows: Suppose the equilibrium allocation fails the CK criterion. Fix the corresponding off-path mechanism m'. We know by hypothesis that the sets $\Xi_1(m')$ and $\Xi_2(m')$ are non-empty. Consider a hypothetical world in which the agent's belief about the principal, say $\tilde{\beta}$, has only support in $\Xi_1(m')$, i.e., $\tilde{\beta} \in B_1(m')$. Suppose furthermore that the set of collusive side mechanisms from which the principal is allowed to propose one, say $\hat{M}(m')$, consists of those mechanisms which induce any $\tilde{k} \notin \Xi_1(m')$ for any $\tilde{\beta} \in B_1(m')$ a payoff weakly lower than equilibrium utility, i.e., $U^p(\tilde{k}) \geq \max M(\tilde{k}, \tilde{\beta}, m')$. The assumption that the equilibrium allocation fails the CK criterion implies that $\hat{M}(m')$ is non-empty, since at least m' is included. We continue by constructing the secured-payoff vector[37] in the hypothetical world, which is characterized by the above stated constraints. The failure of the CK criterion implies that any principal type k with $k \in \Xi_2(m')$ receives secured payoff strictly larger than equilibrium utility. By offering m', k can secure himself for any $\tilde{\beta} \in B_1(m')$ utility strictly larger than his equilibrium utility.

Having established this feature of the secured payoff, we proceed by showing its equivalence to the value of a tractable maximization problem. The solution to this latter problem yields something that can be thought of as RSW allocation in the introduced hypothetical world. In contrast to $(P1)_{\check{k}}$, which defines the original RSW allocation, the latter program is subject to a weakened (PIC) constraint for those types of the principal not in $\Xi_1(m')$. Instead of leaving them with their RSW utility when reporting truthfully, they can garner equilibrium payoff. As final step, we use the constraints of the maximization problem defining the hypothetical RSW allocation to verify that a separation menu exists, if and only if at least one type of the principal strictly prefers the hypothetical RSW allocation to the equilibrium allocation. Thus, the non-existence of a separation menu is sufficient that an equilibrium allocation satisfies the CK criterion.

In order to argue for necessity, let k' be the lowest principal type for which there exists a separation menu. We construct an off-path mechanism, m_{CK}, such that in any continuation equilibria, given any belief about the principal, no type of the principal below k' profits in comparison to equilibrium. To construct m_{CK} we first solve a class of maximization problems, similar to $(P1)_{\check{k}}$. However, the principal's upward adjacent incentive constraints are eased; the menu associated with principal type k, say $\phi_{CK}(\cdot, k)$, satisfies the requirement that principal type $k-1$ prefers either his equilibrium payoff or $\phi_{CK}(\cdot, k-1)$ to $\phi_{CK}(\cdot, k)$. m_{CK} then consists of these K+1 menus. Moreover, the assumption of the existence of a separation menu

[37] The term secured-payoff vector refers to the definition in Mylovanov and Tröger (2015). Inspired by this definition, we define the secured payoff of principal type k the following way:

$$\underline{U}^P_{m'}(k) := \underset{m \in \hat{M}(m')}{sup} \min_{\tilde{\beta} \in B_1(m'), U^P_m \in M(k, \tilde{\beta}, m)} U^P_m(k)$$

where $\hat{M}(m')$ is a set that consists of those side mechanisms, being such that any type $\tilde{k}$ not in $\Xi_1(m')$ receives in any continuation game after the offer of $m \in \hat{M}(m')$ payoff lower than equilibrium utility (for any belief in $B_1(m')$).

implies that at least principal type k' prefers the allocation implemented by m_{CK} to the equilibrium allocation.

To assure that m_{CK} is unanimously accepted by any agent type for any $\tilde{\beta}$, we introduce a lottery that the agent triggers if it is likely, from his perspective, that the mechanism is offered by a principal type below k'. That is, we augment the message space of the agent by a binary report $d \in \{0,1\}$ that he submits together with his type report. If he reports $d = 0$ and the principal submits a report weakly above k', the agent is required to make a small payment. However, he might receive a large payment, $-D^A$, in case the principal reports to have a type below k'. This payment is chosen, such that the agent prefers the collusive side mechanism to the play of the default game for any belief about the principal's types. If the agent reports $d = 1$, then the lottery is degenerate and pays off zero independently of the principal's report.

To assure that the lottery payment is budget feasible and that it does not alter the principal's incentives to report his type truthfully, we make use of money burning, i.e., the mechanism might create a budget surplus. That is, if the agent reports $d = 0$, then the principal has to pay D^A independently of his report. If his report is above k', then this money is burned. Because no type of the principal prefers the agent to report $d = 0$, the above construction assures that an equilibrium does not survive the Intuitive Criterion if there exists a separation menu for at least one type of the principal.

Lemma 2.13. *An allocation fails the CK criterion if and only if for at least one type of the principal there exists a separating menu.*

In order to make use of the condition on the existence of separation menus, the concept of cross subsidizing turns out to be helpful.

Definition 2.13. *Fix principal type k, and a probability $x \in [0,1]$. Define the menu $\phi^x(\cdot,k)$ as the solution to the following problem:*

$$\phi^x(\cdot,k) := \arg\max_{\phi(\cdot,k)} U^p_{\phi(\cdot,k)}(k)$$

subject to:

$$(AIR)^j_k, (AIC)^j_k, (BB)_{j,k} \ \ \forall j \in \Theta^A,$$

$$q^p_{\phi(\cdot,k)} = x.$$

Whenever x is equal to principal type k's equilibrium allotment probability, that is, $x = q^p(k)$, we denote $\phi^{q^p(k)}(\cdot,k)$ by ϕ^E_k and refer to it as principal k's equilibrium menu. Whenever $t^p_{\phi^E_k} \leq (\geq) t^p(k)$ we say that principal type k pays a subsidy (receives a subsidy) in equilibrium.

In words, we fix the probability according to which principal type k receives the right in equilibrium. From his point of view we construct an optimal menu, which implements the same allotment probability under the minimal expected payment to the agent. The menu is subject to the agent's incentive-compatibility and individual rationality constraints, given type k. We say that principal type k is subsidized if the expected transfer payment from the constructed menu is larger than the expected transfer payment in equilibrium.

We employ the above concept to identify sufficient and necessary conditions for the existence of a separation menu for principal type k. To begin with, we show in the appendix that it is without loss of generality to assume that the principal's upward adjacent incentive constraints are satisfied with equality.

Given this, a necessary condition for the non-existence of a separation menu for type k is that he receives a weakly positive subsidy whenever he does not receive the right in equilibrium with certainty. To see this, suppose that k pays a strictly positive subsidy. We construct a separation menu the following way: Starting from k's equilibrium allocation, we marginally increase the allotment probability and the transfer payment such that principal type $k-1$ does not profit from the allocation induced by the constructed menu, in contrast to k. The agent's individual rationality constraints are still satisfied, since k is a subsidizer.

Conversely, one can show that type k has no separating menu if he receives the right with certainty or is cross-subsidized by other types.

As a consequence, any equilibrium that supports efficient collusion fails the CK criterion. Given any such equilibrium allocation, the resulting equilibrium menu of any type of the principal, say k, is common-knowledge menu $\phi^c(\cdot, k)$. This menu can be described the following way: the agent is left with the decision either to buy the right at price θ_k or to select the duty to stay out of the auction at zero compensation. This menu generates the same allotment policy than the equilibrium allocation, and implies payoff for any principal type equal to θ_k. To see that there is at least one type of the principal receiving lower payoff than in equilibrium, i.e., there exists $\check{k}$ such that $U^p_{\phi^e}(\check{k}) < \theta_{\check{k}}$, observe that otherwise every type of the principal would receive larger utility than his type. However, such an allocation is not consistent with both the balanced budget condition and the agent's individual rationality constraints.

Finally, observe that both the RSW allocation and the RSW allocation with entry fee[38] survive the Intuitive Criterion, since any type of the principal who does not receive the right with certainty is (weakly) subsidized.

Proposition 2.6. *An equilibrium allocation survives the CK criterion only if the allotment policy is inefficient in favor of the principal. In particular, the RSW allocation survives the CK criterion. If there does not exist a pool at the top, i.e., $\underline{K} > K$, then the RSW allocation is the only allocation surviving the CK criterion.*

As a consequence of proposition 2.6 the RSW allocation is the unique equilibrium allocation surviving the CK criterion if and only if there does not exists a pool of principal types at the top. This observation is in line with Maskin and Tirole, 1992. The non-existence of the pool at the top implies that their sorting assumption is satisfied. One of their results states that, whenever there is only private information on the principal's side and the sorting assumption is satisfied, then the RSW allocation is the unique equilibrium surviving the CK criterion.

[38] In the case that $\underline{K} < K$ the RSW allocation is such that those principal types in the pool at the top, i.e., in $\{\underline{K}, ..., K\}$, pay a lump-sum transfer to the agent. By the term "RSW allocation with entry fee" we mean the RSW allocation augmented by a lump-sum payment from the agent to the principal, equal to the expected transfer payment the agent receives from those principal types in the pool at the top.

2.5.2 Welfare Implications of the Reserve Price

In the preceding analysis, we focused on equilibrium allocations that survive the Intuitive Criterion. It turned out that if there is no pool at the top, that is, $\underline{K} = K+1$, the RSW allocation is the unique such allocation. If we select the RSW allocation as equilibrium, then a non-trivial reserve price, i.e., a reserve price which excludes a set of types that has strictly positive probability mass from receiving the good, dominates a trivial reserve price in terms of welfare.

Deviating from the approach taken in the preceding analysis, we introduce a reserve price r explicitly and do not allow for positive probability mass on the lowest types of bidders, $\underline{j} = \theta_0 \geq 0$. These lowest types now may be strictly positive, but $F(\underline{j}) = F(\theta_0) = 0$ and $B(0) = \beta_0 \to 0$.

If $\Pi_\phi(j,k)$ is the seller's revenue, given type profile (j,k), the allocation ϕ implies the welfare $W(\phi, r)$, defined by:

$$W(\phi, r) := \mathbb{E}_{j,k}[U^A_\phi(j,k) + U^P_\phi(j,k) + \Pi_\phi(j,k)]$$

$U^A_\phi(j,k)$, $U^P_\phi(j,k)$, $\Pi_\phi(j,k)$, and ϕ depend on r.

Proposition 2.7. *Suppose both bidders' valuation distributions have no point mass at the lowest types, $\underline{j}$ and θ_0, with $\underline{j} = \theta_0 \geq 0$. That is, $F(\underline{j}) = 0$ and $B(0) = \beta_0 \to 0$. If we select the RSW allocation as equilibrium, then a welfare-maximizing seller chooses a non-trivial reserve price, i.e., $r > \theta_0$.*

To see why proposition 2.7 is true, recall that the RSW allocation has the flavor of a least-cost separating equilibrium. That is, it induces the smallest inefficiencies necessary for the principal's types to separate. If the reserve price increases, inefficiencies which result from separation become more costly. Hence, fewer inefficiencies are necessary for separation.

This positive effect of a non-trivial reserve price on welfare must be compared to the well-known welfare decreasing effect of a non-trivial reserve price. In this light, the assumption that both β_0 and $F(\underline{j})$ are arbitrarily close to zero assures that the first effect dominates the latter, and thus a reserve price is desired in terms of welfare.

2.5.3 Bid-Coordination Mechanism

We assume a rather strong kind of collusion technology that commits each player to the outcome of the collusive mechanism. If the mechanism implies inefficient collusion, a player being committed to stay out of the SPA has incentives break the commitment.

We will show that when giving up this assumption, that is, restricting the cartel's collusion technology, the principal can always implement his ex-ante optimal allocation of efficient collusion. Moreover, this allocation also survives the Intuitive Criterion.

In the following we analyze a collusion technology that can only recommend, but not enforce, a cartel member to stay out of the auction. However, we still allow the cartel to exchange side transfers in a budget-balanced manner, and assume that the recommendation mechanism is proposed by bidder P, after having learned the realization of his valuation. Hence, we analyze a setup that is very similar to the so-called bid-coordination mechanism examined in Marshall and Marx, 2007, with the exception that the coordination mechanism is proposed by one of the bidders. Moreover, in order not to rely too much on the all-inclusive cartel assumption, we impose the restriction that no cartel member bids above his valuation.[39] Let us first characterize the (unique) undominated equilibrium allocation and then verify that it can be implemented without enforcement.

We now argue that there exists an undominated equilibrium of the described game, which is consistent with efficient collusion: By the revelation principle, we can focus attention on direct mechanisms, which are such that the bidders behave obediently with respect to the private recommendation. Such a mechanism can induce efficient collusion. For any realized type profile, the highest type of both players is recommended to bid in the auction. As a consequence, a bidder which is recommended to stay out of the auction has no incentive not to behave obediently with respect to this recommendation. Moreover, there cannot exist any other obedient non-trivial recommendation. If the recommendation would imply inefficient collusion, for at least one reported type profile, every bidder has an incentive not to behave obediently when being recommended to stay out of the auction. When breaking the recommendation, there is a strictly positive probability on the event of winning the auction and receiving an additional rent.[40]

Observe that the principal can implement the ex-ante optimal allocation, derived under the enforcement assumption, since he is still able to absorb the entire collusive surplus. That is, to implement the ex-ante optimal allocation no enforcement of the right is necessary. Moreover, because enforcement is not possible, the principal is restricted in his off-path deviations. As a consequence, ϕ^e, satisfies the CK criterion: Recall that an allocation satisfies the CK criterion if there does not exist a separation menu for any type of the principal. Since the equilibrium allocation induces efficient collusion, the existence of a separating menu for principal type k necessarily implies an inefficient high allotment of the right to the principal. Otherwise type k could not separate himself from $k-1$. However, by the arguments above, inefficient collusion cannot be implemented. The agent would not behave

[39] Given the assumption of the all-inclusive cartel, one can implement any equilibrium allocation under the strong collusion technology, i.e., under enforcement, also under the weak technology, i.e., no enforcement, by employing the following recommendation mechanism: After having submitted the transfers, the bidder who is supposed to receive the right to be the only bidder in the auction is advised to submit a very high bid, whereas the other potential bidder is advised to bid 0. See Ungern-Sternberg (1988). Behaving obedient to this recommendation is a weakly dominated action for the bidder that is supposed to submit the high bid.

[40] In this game, the characterization of equilibrium is straightforward: Every feasible mechanism that induces each principal type a payoff weakly larger than the one generated by the non-cooperative play of the SPA can be supported as equilibrium. To see this, consider the following off-path belief of the agent on the principal: The principal is the lowest type 0. Given this belief, the agent always plays the SPA non-cooperatively, independent of the recommendation. Given this continuation strategy, budget balance implies that the mechanism turns into a zero-sum game with type-independent strategies. Since the off-path belief is common knowledge, the principal can assert that all agent types bid their valuation in the SPA. As a consequence, every principal type only cares about the transfer payment and not about the recommendation generated by the mechanism. The agent is therefore only willing to participate in the game if the expected payment is weakly positive. In this instance, every principal type receives weakly less than his non-cooperative outside option. Yet, by offering a degenerate mechanism that specifies zero transfers, the principal can always secure himself the non-cooperative SPA payoff. Thus, any efficient allocation that leaves every type of both bidders with weakly larger payoff than the non-cooperative play of the SPA can be supported in equilibrium.

obediently with respect to a recommendation inducing inefficient collusion and thus such inefficient allotment is not feasible.

The self-enforcing collusive agreement can be implemented by means of a linear-separable transfer scheme. That is, first the agent pays a participation fee to the principal equal to the height of the expected type-dependent transfer the principal has to pay, conditional that collusion is efficient. Afterwards, both the principal and the agent simultaneously submit to each other the type-dependent part of the transfer scheme, where the lowest types pay a transfer of zero. These transfers perfectly reveal the type of the player submitting it, and thus, staying out of the auction is a best response for the player with the lower valuation draw.[41] At no stage does a bidder have an incentive to break the collusive agreement. If, for example, the principal collects the agent's entry fee but does not submit a transfer payment himself, this is equivalent to submitting the transfer payment of the lowest principal type, which is not desired by incentive compatibility.

2.5.4 Non-All-Inclusive Cartel

Suppose we generalize our set-up slightly by introducing a third bidder, not a member of the cartel that bids according to his weakly dominant strategy. We suggest that all of the analysis is still valid under minor changes of notation. That is, we replace the type of a cartel member by his interim expected net utility of playing the SPA non-cooperatively against the outside bidder.[42] Given this modification, in an equilibrium that survives the Intuitive Criterion inefficiencies with respect to the cartel members due to collusion are still present. In what follows we focus entirely on the inefficient RSW allocation.

If the cartel is not all-inclusive, the characteristics of collusion have implications on the seller's revenue. Since the member that receives the right to be the cartel's only bidder in the auction submits a bid equal to his valuation, the seller prefers collusion to be efficient. This is the case because efficient collusion is the form of collusion that maximizes the cartel's expected bid submitted in the SPA.

If the cartel is not all-inclusive and collusion is efficient, it is well-known from the literature that a revenue maximizing seller can subsume those bidders inside of the cartel under a single bidder, the cartel bidder, with valuation equal to the maximum valuation of the cartel's members. The seller then sets the reserve price

[41]One might wonder how decent it is to assume the play of such a weakly-dominated actions. However, note, if one assumes that both the agent and the principal have discrete types and there are small bidding costs, smaller than the distance between any two adjacent types, all of the presented arguments are still valid, under minor changes. Still, submitting a bid of zero is not a dominated action.

[42]To be more specific, we transform the type of a cartel member, say j, into the valuation of the right $v(j)$. If the outside bidder's type is distributed continuously according to $F^o(\cdot)$, with support equal to the one of the agent's distribution, then e.g., agent type j's valuation of the right is given by $v(j) = \int_r^j F^o(v)dv$. Moreover, to calculate, for example, the RSW allocation one first must determine the utility the agent receives when rejecting the mechanism, given the principal is type k, with $j > \theta_k \geq r$ (otherwise agent j receives utility zero). This object takes the form $U_D^A(j,k) = \int_{\theta_k}^j F^o(v)dv$. With this, one still can think of the RSW allocation as being induced by price and compensation pairs the principal offers to the agent. Suppose the principal's type is k, and the threshold agent type, being indifferent between paying the price and picking the compensation, is $j(k)$. Then the compensation, c, takes the form $c = U_D^A(j(k),k)$ and the price, $\tilde{p}$, is $\tilde{p} = v(j(k)) - U_D^A(j(k),k)$. One can then find the RSW allocation by solving a sequence of problems, similar to $\{(P1)_k\}_{k=0}^K$ but modified in the manner described above.

as if the cartel bidder, with distribution as described above, plays the auction non-cooperatively against the outside bidders.

In contrast, if we focus on an inefficient equilibrium, the seller's problem differs in two respects. First, the distribution of the cartel player changes, since collusion is not efficient. Second, the distribution of the cartel bidder might depend on the reserve price.

Concerning the implications on the seller's optimal reserve price, note that both effects work in favor of a larger reserve price, in comparison to the situation in which collusion is organized by a third party and therefore is efficient. First, if collusion is inefficient, then the distribution of the second-highest bid that is submitted in the SPA, being played by the cartel bidder against the outside bidder, is first-order stochastically dominated by the respective distribution if collusion is efficient. However, given any reserve price, the probability of the event that the highest bid being submitted in the SPA is above the reserve price is not affected by the characteristics of collusion, as according to the RSW allocation, the right to bid is either allotted to the agent or to the principal if at least one of both players has a valuation above the reserve price. Employing the reasoning of Krishna (2002), section 11.1.2 on bidding rings, this effect in isolation already implies a larger reserve price when collusion is inefficient compared to the case of efficient collusion. Second, as an increase of the reserve price decreases the inefficiencies inside of the cartel, and in turn increases the expected bid of the cartel bidder, there is a second channel that increases the optimal reserve price.

2.6 Discussion of the Number of Cartel Members

An undeniable limitation of our model is the number of bidders. In the SPA, the all-inclusive cartel assumption is without loss of generality (see section 2.5.4). However, in the FPA, a non-all-inclusive cartel implies that the agents' outside options are determined by an asymmetric FPA not featuring analytic solutions in general. Admittedly, this is very unsatisfying. Yet, the entire bidder collusion literature suffers from this annoyance.[43]

In the following, we state sophisticated **conjectures** concerning the implications of increasing the number of cartel members from 2 to $N+1$. In the FPA, we strongly believe that the following happens: As strategic player, the principal implements a no-veto constraint mechanism. That is, if an agent rejects the collusive side mechanism the remaining cartel members collude according to some predefined protocol. To punish the deviating agent the most, - recall, the collusive side mechanism can only allot the right to bid -, the remaining cartel members collude efficiently. As a consequence, an agent's outside option is determined by the asymmetric FPA being played between him and the cartel member with the highest valuation, with degenerate belief concentrated on the largest type of the former. Given this, one can adjust the mechanism $\tilde{m}$ to $\tilde{m}^a$ such that the latter replicates the non-cooperative

[43]Papers that deal with this problem have to rely on particular examples, often featuring only numerical solutions (see McAfee and McMillan, 1992, Marshall et al., 1994 and Marshall and Marx, 2007).

play of the described setting.[44] A similar reasoning as in the one-agent case should show that $\tilde{m}^a$ implements the principal's secured payoff and the proposal of $\tilde{m}^a$ can be supported as equilibrium of the mechanism-selection game. However, establishing uniqueness is a more elaborate task. This involves to show that $\tilde{m}^a$ implements the highest principal type the largest feasible payoff given the prior beliefs about the principal. This is essentially a mechanism design problem, involving the careful specification of measures that trade off any agent's type-depending individual rationality constraints. If there is more than one agent, the construction used in the proof of theorem 2.11 does not directly go through. Instead, a more elaborate construction of measures as in the flavor of the proof of lemma 2.8 is needed. We are not entirely certain whether this is feasible, that is, whether there is indeed a unique equilibrium.

One might wonder why there are no equilibria in which the agents implicitly collude against the principal and coordinately reject certain mechanisms that leave them with too low payoffs. That is, why doesn't there arise a veto-constraint mechanism ? The reason is that the principal can prevent the above-described collusive behavior of the agents. He can specify large payments for those agents that accepted the mechanism, conditional on the event of at least one agent rejecting the mechanism. By this, the event that all agents reject the mechanism becomes an off-path event, and the relevant individual rationality constraints are those being described in the previous paragraph. Because of this, veto-constraint mechanisms only arise if one restricts the principal's set of side mechanisms to those. It is actually *not at all* an interesting question to explore to what extent the agents can make use of the above collusive behavior to influence the principal's proposal to their advantage. Restricting the principal's set of mechanisms to those that are veto-constrained implies that basically every proposal can be supported as equilibrium of the mechanism-selection game. Given that the agents have decided to support mechanism m', they coordinate on the following strategy: Accept mechanism m', and if the principal proposes a different mechanism ,reject it. Given this strategy, at most mechanism m' can be supported in equilibrium. It can be supported if there exist certain off-path beliefs about the agent and the principal such that every bidder receives in the non-cooperative FPA, played with this certain off-path belief about himself and prior belief about the other bidders, less payoff than from the play of m'. Hence, if the set of side mechanisms were restricted too much, the model would fail to capture the strategic interaction inside of the cartel.

The above remark about veto-constraint side mechanisms also applies to the SPA. Focusing on no-veto-constraint mechanisms, there are no additional insights from increasing the cartel's size. The cartel members' interaction is governed by the principal's signaling motive, which does not rely on the number of agents. However, with more than one agent, the principal receives a rent from organizing collusion between the agents. As a consequence, if the number of agents increases, the inefficiencies induced by the principal's signaling motive quantitatively decrease.[45]

[44]Suppose there are N agents. Index a generic agent by i. Let $H^{-i}(t)$ be the distribution of the maximum valuation, t, of the cartel without agent i. Each member reports in $\tilde{m}^a$ a number $t^i(j) := \arg max_t H^{-i}(t)(j-t)$. The allocation is then such that principal type k receives the right if $\theta_k > \max_{i\in\{0,\ldots,N\}} t^i(j)$, and else the agent i that reported the largest $t^{\cdot}(j)$.

[45]Indeed, concerning the SPA ,let us focus on the characterization of the RSW allocation. With more than one agent, all being ex-ante symmetric, this allocation takes the following form: Depending on the principal's type, there is a threshold. An agent receives the right if and only if he has the

2.7 Conclusion and Discussion

The aim of this paper was to analyze the implications of the auction format on the characteristics of bidder collusion. To suppress competition, bidders form a cartel and organize their joint bidding behavior. The key premise in our analysis is that it is the cartel members themselves who resolve the information asymmetry among them. This necessity gives rise to strategic interactions.[46] We model this interaction by giving one cartel member the right to act as principal and to propose the other member, the agent, a collusive side mechanism, whose play determines the enforceable collusive allocation.

One contribution of this paper is to find that collusion leads in general to inefficiencies, both in the SPA and in the FPA. This novel result stands in contrast to the literature on bidding rings, which abstracts from the signaling motive and predicts the cartel to collude efficiently in both auction formats. By this, the cartel can achieve its joint profits-maximizing benchmark.

Focusing on the SPA, we find that the cartel's interaction is entirely driven by the principal's signaling motive. It is shown that high principal types face a tradeoff between maximizing and extracting the cartel's expected collusive profits and proposing inefficient allocations in order to signal their strength in exchange for a better bargaining position inside of the cartel. Loosely speaking, if both bidders are not too asymmetric or the stronger of them becomes the principal, efficient collusion can be supported in equilibrium. With this result, we establish that the cartel's joint profit-maximizing benchmark of efficient collusion is robust to a mild form of signaling.

In contrast, we find that the cartel cannot achieve this benchmark in the FPA where collusion is always inefficient. Importantly, the non-cooperative play of the FPA depends on the belief the bidders hold about each other. By this, his belief about the principal affects the agent's inference that he is supposed to draw from the principal's mechanism proposal. Making use of his large mechanism space, the principal exploits this bid-shading property of the FPA by proposing a mechanism that replicates the non-cooperative play of the auction for any belief the agent may hold about him. Given this mechanism, the cartel's interaction is driven by the agent's fear that the breakdown of collusion signals his strength, which results in a distorted play of the FPA to the agent's disfavor.[47]

---

highest valuation above the threshold. Otherwise the principal receives the right. Note that even if the agents are ex-ante asymmetric, introducing further inefficiencies is costly in terms of their binding individual rationality constraints. As a consequence, we suspect that the principal treats all of them equally in case they have valuation above the threshold. Technically, because the agents' outside options are increasing in their types, the respective individual rationality constraints are allowed to bind at the optimum for all types above the threshold. Given this, one should be able to construct measures such that the agents' virtual valuations are equated. That is, everything the principal saves on one agent's (AIR) constraint by introducing inefficiencies w.r.t. the agents implies costs of at least the same amount on some other agent's (AIR) constraint.

[46] This is a reasonable premise. E.g., Cave and Salant, 1987 document cases of legal cartels in agricultural markets. The interaction between the members resulting from their need to agree on quotas was driven by non-cooperative, strategic forces.

[47] In an experiment Llorente-Saguer and Zultan, 2014 tested Eső and Schummer, 2004 / Rachmilevitch, 2013's take-it-or-leave-it bribing scheme both in the FPA and SPA. They find that collusion in the FPA does not break down and leads to inefficiencies. These inefficiencies are due to unsuccessful collusive attempts, which distort the non-cooperative play of the FPA. This is in line with our results. We find a close connection between the collusive allocation and the allocation induced by

Our paper highlights that the channels that govern the cartel's interaction depend on the auction format. Nonetheless, we are still able to draw some quantitative comparisons between collusion in the SPA and the FPA. We find that the least inefficient equilibrium in the SPA is less inefficient than the unique equilibrium of the FPA.

Naturally, our results depend on the assumed bargaining protocol inside of the cartel. We faced the following modeling challenge: Our aim was to capture the strategic interaction between the cartel members without predetermining the collusive side mechanism resulting from this interaction. This latter point allows us to meaningfully relate our results to well-established findings on bidding rings. Similar to this literature, we took a mechanism-design approach with the important difference that a member inside the cartel proposes the mechanism. We thus assumed an extreme distribution of bargaining power and solved a non-standard informed-principal problem.

In our point of view, the assumption of the extreme distribution of bargaining power is rather innocuous, as any distribution can be restored by accordingly randomizing which of the bidders becomes the principal. However, our conclusions concerning the FPA might appear to depend on the large set of side mechanisms available to the principal. The unique equilibrium heavily relies on the fact that the principal can propose a mechanism that is not a direct revelation mechanism, replicating the non-cooperative play of the auction for any belief about him.

Be that as it may, this unexpected result, - in fact, this is the very first result for an informed-principal problem in a setting where outside options are determined by the play of a belief-dependent game - , highlights an economically relevant channel; by influencing each cartel member's inference, drawn from interacting with the other members, the FPA makes it impossible for the cartel to achieve efficient collusion.

This insight draws attention to a direction for future research. Anti-collusive auction design typically takes the cartel's interaction as given and focuses on design elements, making it hard for the cartel to enforce the collusive agreement (e.g., see Marshall and Marx, 2009). Importantly, such design elements (for example allowing for shill bidders and not revealing the identity of the auction's winning bidder) are redundant if the auction is played non-cooperatively, but imply anti-collusive gains if a bidder cartel is present. However, there are environments where the identity of the winner cannot be kept secret, like procurement auctions of construction services. Deterring the cartel from enforcing its agreement is hard, if not impossible. Exploring whether there are auction-design elements that influence the cartel's interaction in an anti-collusive way seems to be relevant.

---

the non-cooperative play that would result if the agent were to reject the collusive side mechanism. We differ from the experiment in that the principal is not restricted to bribes and therefore proposes a mechanism that replicates the allocation being induced after an *hypothetical* unsuccessful collusive attempt.

Chapter 3

Managing a Conflict: Alternative Dispute Resolutions in Contests

with Johannes Schneider

3.1 Introduction

Alternative Dispute Resolution (ADR) is a tool introduced into the legal system of many countries to increase the system's efficiency by settling as many cases as possible outside court. ADR itself can take many forms and describes a third-party mechanism other than formal litigation to solve the conflict. However, ADR typically cannot overturn the rule of law, such that parties return to the litigation track once ADR fails. Given that ADR and litigation remain thus connected, several questions arise. How does the information exchanged during ADR influence the behavior in litigation post ADR-breakdown? How does the threat of ADR-breakdown influence the litigants' willingness to release information during ADR? How should we design ADR "in the shadow of the court"?

The aim of this paper is to study the optimal third-party ADR-mechanism that uses litigation as the fall-back option in case no agreement is reached. We provide a model identifying the two-way channel that links an optimal mechanism (ADR) and an underlying contest (litigation). We show that optimal ADR and litigation cannot be considered as independent problems: the information revealed in the ADR-stage influences the choice of action in both ADR and litigation. Litigants' investment into evidence provision after breakdown depends on the beliefs about their opponent's action. The ADR-designer needs to be concerned about managing the players' beliefs in case ADR breaks down. Moreover, ADR cannot fully eliminate litigation as parties differ in their marginal cost of evidence provision. ADR breaks down sometimes to screen parties and to ensure truth-telling during ADR.

Most modern societies accept the concept of the "rule of law" despite an overburdened legal system: in 2014 each judge in the U.S. district courts received 658 new cases. At the same time the number of pending cases is even larger with 694 per judge. The large caseload leads to a median time from filing to trial of around 2 years. As litigation requires a lot of time and resources from courts, each case that forgoes litigation also has a positive externality on the functioning of the legal system as a whole.

Thus, most jurisdictions encourage parties to engage in some form of ADR before starting the formal litigation process. The U.S. Alternative Dispute Resolution Act of 1998 states that courts should provide litigants with ADR-options in all civil cases. ADR is defined as "any process or procedure, other than an adjudication by a presiding judge, in which a neutral third party participates to assist in the resolution of issues in controversy" (Alternative Dispute Resolution Act, 1998). However, ADR supplements the "rule of law" rather than replacing it. Ultimately, each party has the right to return to formal litigation.[1] Hence, ADR indeed happens "in the shadow of the court:" whenever no settlement is achieved via ADR, litigants return to the traditional litigation path.

Nonetheless, ADR is a very effective tool to settle conflicts and has success rates substantially above 50% across time, jurisdictions, and case characteristics. Furthermore, litigants report that ADR has an impact on the continuation of the trial even if unsuccessful (Genn, 1998; Anderson and Pi, 2004). The informational spillovers to post-breakdown litigation influences the design of optimal ADR: if the information a player receives during ADR depends on the information she provides, parties have an incentive to strategically extract information *within ADR* which they can use *in litigation* once ADR breaks down.

We follow a large literature dating back to Posner (1973) and consider litigation as a legal contest (for an overview on the litigation literature see Spier (2007)). The party providing the most convincing evidence wins the case. In such a contest, the optimal amount of evidence the plaintiff provides is a function not only of her own cost of evidence provision, but also of her beliefs about the defendant's evidence choice and vice versa. Hence, litigation strategies after ADR-breakdown are a function of the players' *belief system*.

Optimal ADR-design should take the belief-channel into account to ensure incentive compatibility: suppose a plaintiff who only has access to circumstantial evidence reports to the mediator instead that she has direct evidence. She then might gain from misreporting in two dimensions. First, through a direct effect: reporting better evidence can lead to a more favorable settlement. Second, there is an indirect effect: if the plaintiff misreports, she may also benefit if ADR fails to resolve the conflict. By misreporting in the ADR stage, the plaintiff may influence her post-breakdown expectation about the defendant's type since breakdown is a function of both players' reports. Changing the beliefs post-breakdown affects expected litigation outcomes and provides an additional incentive to misreport. While the direct effect is present in standard mechanism design models, we seem to be the first to consider the indirect effect as the outside-option of our mechanism depends on the belief system.

Our analysis highlights several important features of ADR in the shadow of a legal contest: we show that if ADR cannot promise full-settlement for *all* type-profiles, then ADR cannot promise full-settlement for *any* type-profile. The reason is that if the mediator promises settlement for a specific type-profile, it imposes an externality on the other types by influencing their breakdown beliefs.

We further show that the optimal mechanism is always asymmetric. It favors one player when ADR breaks down and the other when ADR is successful, even when players are fully symmetric ex-ante. At the time of participating, players only care about their expected valuation being the sum of the valuations in case of both settlement and breakdown. To keep the expected valuations constant, the valuation promised to players in settlement must increase the more competitive and therefore wasteful litigation post ADR-breakdown is. Consequently, optimal mediation

[1] For a detailed discussion on this, see Brown, Cervenak, and Fairman (1998).

makes the litigation process post-ADR less competitive by inducing asymmetric beliefs to save on resources needed for settlement.

While the optimal mechanism results in asymmetric beliefs, it ensures that beliefs are independent of the player's type-report. If a player could obtain different information from different reports, she could induce a situation without common knowledge of beliefs post-breakdown: the deviating player knows that she misreported, but her opponent does not. Each player's optimal action depends on both her own belief about the opponent and what the opponent thinks this belief is. Learning from reports can thus provide an incentive to misreport in hope of breakdown. If beliefs are independent of the report, however, such a problem does not arise because deviations do not create an information advantage.

We significantly differ from standard models of conflict resolution in that we consider a model in which investment into the conflict is made *after* the resolution mechanism broke down. Nonetheless, a key result derived by Hörner, Morelli, and Squintani (2015) carries over to our setting: if the mediator can talk to parties in private, the players' level of commitment is not important. Compared to a situation in which parties commit to the mechanism at an interim stage, the mediator can achieve (almost) the same result if parties are allowed to unilaterally opt-out of mediation after the settlement proposal. The reason is that private communication allows the mediator to conceals some information even at an ex-post stage.

Our findings contribute to the ongoing discussion of optimal ADR-design by pointing out several important aspects: (1) optimal ADR can settle most of the cases outside court independent of the cases' characteristics; (2) the level of commitment needed by the parties is not important if the mediator can communicate to parties in private; (3) regulators should be careful when preventing mediators from using asymmetric protocols as they increase the probability of ADR breaking down; and (4) to incentivize settlement, optimal ADR should predominantly manage beliefs in case a breakdown occurs.

We also contribute to the literature on mechanism design. If screening can happen only through an underlying game, on-path breakdown is informative for players and necessary for optimality. Our model emphasizes the relevance of belief management by the mechanism if the underlying game, and thus the outside option, is belief dependent. Our findings directly apply to other situations in which a wasteful contest is the last resort such as strikes, political lobbying, patent races, and standard setting organizations.

Outline. After discussing the literature in Section 3.2, we set up the model in Section 3.3 and derive the optimal mechanism in Section 3.4. Subsequently, we discuss the findings in Section 3.5 and several extensions in Section 3.6. Section 3.7 concludes.

3.2 Related Literature

We contribute to three strands of literature: (1) to the best of our knowledge we provide the first formal model in the law and economics literature that explicitly addresses the complementarity of litigation and ADR; (2) we add a new channel to the literature on mechanism design with endogenous outside option by showing that a mechanism which cannot fully avoid a post-mechanism game should be concerned about the information release during the process; and (3) we add to the existing literature of mechanism design and conflict resolution as we consider a setup in which parties make their decision on investment into the default game *after* the conflict arises.

We connect to the law and economics literature on settlement under asymmetric information dating back to the seminal paper by Bebchuk (1984). Spier (1994) is the first in this line to consider a mechanism design approach. She uses a model that applies to situations in which investment in evidence provision was made *prior* to negotiations and is interested in optimal fee-shifting between parties. We differ in two aspects: we hold the rules of litigation fixed and study a model in which the choice on how much evidence to present is made *after* settlement negotiations. This results in an optimal mechanism that conditions on informational spill-overs of ADR onto litigation.[2]

Brown and Ayres (1994) highlight that managing the information flow between litigants can be a rationale for ADR that goes beyond reducing psychological barriers to negotiation. There is, however, to the best of our knowledge no paper yet, that links information exchange in pre-litigation ADR with litigation as a strategic game. We model litigation in the tradition of Posner (1973) as a legal contest.[3] Our findings show that such a link is important as ADR and litigation should not be treated as two independent problems, but two stages of the same game.

The second strand of literature we relate to is that of mechanism design with endogenous outside options, i.e. mechanisms which cannot fully replace an underlying strategic game. Similar to Cramton and Palfrey (1995), the optimal ADR mechanism is ratified by both parties. Without mutual consent, parties play the litigation game. However, in our model, mediation sometimes breaks down after participation and parties are referred to the underlying game. Celik and Peters (2011) show that for some games it is optimal to design a mechanism without full participation. In our model, this channel is not present and full participation is optimal. Instead, we explore an additional channel: we ask how on-path references to the default game *by the mechanism* interact with the belief structure of the players after breakdown.

We also connect to the literature on conflict resolution as the two closest papers to ours are Bester and Wärneryd (2006) and Hörner, Morelli, and Squintani (2015). Bester and Wärneryd (2006) were the first to study conflict resolution in a mechanism design environment. Similar to us, they look for the conflict minimizing mechanism and find that it is typically stochastic. Hörner, Morelli, and Squintani (2015), building on Bester and Wärneryd (2006), study optimal mediation in the context of international relations. They show that limited commitment of the disputants does not change the outcome of the optimal mechanism as long as the mediator can talk to parties in private.

The main difference between our model and those of Hörner, Morelli, and Squintani (2015) and Bester and Wärneryd (2006) is the timing of events: through their fixed, type-dependent outside option, Hörner, Morelli, and Squintani (2015) implicitly assume that investment decisions take place *before* the conflict arises. While

[2] Another recent paper discussing third-party mediation is Doornik (2014) who studies the optimal use of a fixed mediation mechanism. Different from us, she is interested in *when to use* a certain ADR mechanism,while we focus on *the optimal design* of ADR.

[3] Examples include Baye, Kovenock, and Vries (2005), Prescott, Spier, and Yoon (2014), Spier and Rosenberg (2011), and Katz (1988). In addition, see Spier (2007) for a general discussion on litigation in the law and economics literature.

this assumption may apply to mediation attempts in international relations, it applies less to ADR negotiations as the collection of evidence typically happens *after* the conflict arises. Our results are thus a complement to Meirowitz et al. (2015) who study the relationship between dispute resolution and *pre-conflict investment*. Contrary to that, we study the relationship between dispute resolution and *post-mediation investment*. An important result of Hörner, Morelli, and Squintani (2015), however, carries over to our setting: limited commitment changes the result of the optimal mechanism arbitrarily little.

Although the result on limited commitment is similar, the optimal mechanism itself is qualitatively different: in Hörner, Morelli, and Squintani (2015), the result is always symmetric and involves full-settlement between weak types. In our setup neither occurs: the optimal mechanism is never symmetric and mediation has a positive breakdown probability for all type profiles as weak types are needed in the post-mediation contests to ensure full participation which is always optimal.

Our concept of mediation is based on Bester and Wärneryd (2006) and lies between pure communication devices as in Mitusch and Strausz (2005) and a mediator with independent sources of information (Fey and Ramsay, 2010). Pavlov (2013) shows that the former has no effect on the outcome in contests but, different to Fey and Ramsay (2010), the mediator can resolve the majority of conflicts without the need of an exogenous information source.

3.3 Model

Litigation Game. The underlying litigation game Γ of our model is an all-pay contest with asymmetric information as in Szech (2011) and Siegel (2014).[4] There are two risk-neutral players $i = 1, 2$ who compete for a good of a commonly known value of 1. Both players simultaneously decide on a score s_i and the player with the highest score wins the good. Ties are broken in favor of player 1.[5] Obtaining a score is costly. Players are ex-ante symmetric and have low marginal cost, c_l, with probability p, or high marginal cost, $c_h \equiv \kappa\, c_l$; $\kappa > 1$, with probability $(1 - p)$. All but the realization of the cost, which is privately learned by each player, is common knowledge. To simplify notation, we denote the low-cost type "l" and the high-cost type "h". In line with this simplification, we are going to use the expressions "player i, type k" and "player ik" interchangeably.

Mediator. We model the mediator as a neutral third-party possessing no private information who announces a protocol $\mathcal{X}$ and has the ability to commit to it. The protocol is a mapping from a message profile, M, to triple (G, X_1, X_2) where G denotes the matrix of breakdown probabilities and X_i the matrix of settlement shares. Mediation is voluntary. If a player refuses to participate in the mechanism, this event becomes commonly-known and litigation is played. A result of (Celik and Peters, 2011) implies that we can focus without loss of generality on mechanisms

[4] We follow the terminology of Siegel (2009), indicating that players have heterogeneous cost of effort but a common perception of the prize.

[5] This technical assumption allows us to circumvent openness problems off-path. However, any other tie-breaking rule would work at cost of additional notation.

that induce full participation.[6] Suppose for the moment that the mediator proposes a direct-revelation mechanisms. Thus, let

$$G = \begin{pmatrix} \gamma(l,l) & \gamma(l,h) \\ \gamma(h,l) & \gamma(h,h) \end{pmatrix},$$

and

$$X_i = \begin{pmatrix} x_i(l,l) & x_i(l,h) \\ x_i(h,l) & x_i(h,h) \end{pmatrix},$$

where $\gamma(M)$ denotes the probability of mediation breakdown after message profile $M = (m_1, m_2)$, that is the probability that players are sent back to the litigation game Γ after message M. Further, $x_i(M)$ denotes the share of the good assigned to player i after M.[7]

We assume budget balance and non-negative shares: the designer can only divide the good in question and allocate shares to players. These shares sum up to not more than one, that is $x_1(k_1, k_2) + x_2(k_1, k_2) \leq 1$.[8]

The revelation principle implies that the outcome of any equilibrium being induced by some ADR mechanism, can be replicated by a direct ADR mechanism in which the mediator sends action recommendations to the players, conditional on breakdown. We restrict our attention to the following class of ADR games: Take any continuation game that begins after the breakdown of ADR with the players choice about their score in the litigation game. If this continuation game is part of the equilibrium path of the grand game (i.e., the game that begins with the player's decision whether to participate in the ADR game), then players' (type-dependent) beliefs about payoff types are commonly known.[9,10] Observe, this class of games includes, in particular, any ADR process in which the mediator publicly speaks to the players and also any decentralized form of ADR, in which the players directly communicate with each other. In principle, we could allow the mediator to send recommendations to the players that are obedient and are such that the players' inference, drawn from the received recommendations given a commonly-known recommendation protocol, are consistent with common-knowledge of beliefs with respect to payoff-types.[11] Yet, to characterize the optimal solution, we can abstract

[6]If one player rejects the mechanism, his opponent may update his belief. Since the priors are type-independent, the posteriors are, too. We verify in the preceding analysis that the payoffs induced by the litigation game are weakly convex in beliefs, if beliefs are type-independent. By (Celik and Peters, 2011) this feature assures that full-participation is without loss of generality. See lemma B.1.

[7]For the ease of notation, we assume without loss of generality that the message k is assigned to the meaning "I am type k".

[8]If the good itself was indivisible, a lottery could implement the same result.

[9]Without this restriction, the mediator could in principle induce a communication equilibrium in the litigation game. Pavlov (2013) shows that all communication equilibria in all-pay contests are payoff equivalent to the unique Nash equilibrium. Yet, Pavlov (2013)'s setting is not identical to ours and his proof does not directly apply. Although we strongly conjecture that the imposed restriction is without loss of generality, a formal proof is still required.

[10]A more natural restriction is to require common-knowledge about payoff-types beliefs to hold also off the equilibrium path. The solution we derive also satisfies this more demanding restriction.

[11]I.e. recommendations which can be implemented by the play of a Bayes Nash equilibrium, possibly augmented by a randomization device, given some commonly-known beliefs about payoff-types.

from such recommendations or signals.[12]

We are looking for a mechanism that minimizes the ex-ante probability of mediation breakdown, $Pr(\Gamma)$. The solution concept is perfect Bayesian equilibrium.

Timing. For most of the analysis, we consider an interim individually rational mechanism.[13] Hence, the timing is as follows: first, the mediator commits to the mediation protocol $\mathcal{X}$ and players learn their type privately. Second, players simultaneously decide whether to participate in the mediation mechanism. If any player rejects, players update beliefs and play the litigation game. If both accept, players privately send a message m_i to the mediator.

Following her protocol $\mathcal{X}$, the mediator either implements an allocation (x_1, x_2) or initiates breakdown. In the latter case players update beliefs and go to litigation.

Discussion of the Assumptions. We follow a large strand of the literature in assuming that litigation is a legal contest. The all-pay contest, a limiting case of a general Tullock (1980) contest, is only assumed to ensure closed form solutions.

As expected, contest utilities are continuous for every action pair, hence adding noise would not change our results qualitatively.[14] The same is true for the constant marginal cost of evidence production. Results maintain if we assume a more sophisticated (monotonic) evidence provision function as used e.g. in Baye, Kovenock, and Vries (2005). Ex-ante symmetry is chosen for simplicity, too, and can be relaxed without changing the results.

The assumption that mediation is designed by a neutral third-party follows the U.S. Alternative Dispute Resolution Act of 1998. In practice, ADR is typically conducted by (retired) judges, law professors or private mediation companies all repeating the mediation services on a regular basis. Clearly, trust is a relevant issue for those mediators and provides a rationale for commitment.

Interim individual rationality of the players is assumed for the ease of notation, only. In Section 3.6 we show in line with the argument by Hörner, Morelli, and Squintani (2015) that assuming ex-post individual rationality changes results only arbitrarily little.

Finally, the assumption that the mediator aims to minimize breakdown is in line with the theoretical literature on conflict resolution. Courts have an enormous backlog in pending cases. Mainly because of the backlog, the time from filing to trial takes typically more than two years. Decreasing the number of court cases therefore has a positive effect on caseloads as well as on possible future conflicting parties and their ability to use the legal system effectively. Related to that, reducing the backlog is the main goal of ADR in practice: the success of dispute resolution programs is typically measured in the share of cases settled (see, e.g., Anderson and Pi (2004) and Genn (1998)). Moreover, the assumption that ADR minimizes the number of court cases adds to the tractability of the model: contest utilities are

[12] To see this, suppose the mediator releases public signals in the litigation game, where each public signal gives rise to a different system of first-order beliefs, i.e. beliefs over payoff-types. Each such signal induces a different equilibrium play of the litigation game. Hence, the mediator optimally implements the signal that induces his most preferred litigation play with probability 1. The proof can be found in appendix (lemma B.5).

[13] In Section 3.6 we show in an extension that assuming ex-post individual rationally can changes results arbitrarily little.

[14] See e.g. Baye, Kovenock, and Vries (1996), Ewerhart (2015), and Che and Gale (2000) for a detailed discussion.

not well behaved in the mediator's choices. A different objective complicates the analysis substantially by adding non-convexities to the objective function.

3.4 Analysis

We proceed with the analysis in several steps. First, we characterize the equilibrium of the continuation game after on-path breakdown for a given information structure. Next, we characterize the properties of the continuation game following a misreport during the reporting stage. Breakdown after a false report essentially produces a situation without common knowledge of beliefs and provides the deviator with an informational advantage. We show that all players and types weakly prefer the on-path contest to the deviation contest only if beliefs are independent of their type reports. The third step is to rewrite the problem to overcome non-convexities and to make it tractable. Litigation is the only source of screening, and thus, the mediator is concerned about choosing the optimal information structure post-breakdown. This determines the solution of the problem up to a constant. We show that this constant is entirely determined by the fact that the optimal mechanism is budget balanced. Finally, we characterize the optimal mechanism. We show that it discriminates even between symmetric players, but involves a type-independent belief structure.

We organize the remainder of this section as follows: for each step we first state its result and provide an intuition thereafter. Formal proofs are provided in Appendix B.3.

3.4.1 Equilibrium Characterization of the Continuation Game

The continuation game after breakdown of mediation is an all-pay contest with type-dependent probabilities as defined in Section 3.3.

Let $p_i(k_i|m_{-i})$ denote the probability that *player i is of type* k_i, given that *player -i is of type* m_{-i}. For readability, we drop the player subscript in the arguments and write $p_i(k|m)$. In contests, the literature typically assumes some form of monotonicity condition which guarantees that having a low-cost type is desirable for all players. We follow Siegel (2014) and call the environment monotone if

$$\frac{p_i(k|l)}{p_i(k|h)} > \frac{c_l}{c_h} = \frac{1}{\kappa} \qquad \forall\, i, k. \tag{M}$$

In what follows, we are going to assume that (M) holds, i.e. we assume that it is optimal for the mediator to induce post-breakdown belief structures that satisfy (M). In the Appendix we show that this is indeed optimal even if the mediator could choose non-monotone environments.[15] Further, we assume throughout the paper that the probability that player 1 has low-cost, given player 2 reported low-cost, is weakly larger than the probability that player 2 has low-cost, given player 1 reported low-cost. Hence, player 1 is the stronger player in the contest or $p_1(l|l) \geq p_2(l|l)$. This assumption is without loss of generality.

[15] Siegel (2014) shows that in principle little can be said if (M) is violated. In our setting, the mediator can only induces Bayes' plausible belief structures. Thus, it is actually possible to characterize the non-monotonic equilibria explicitly. We characterize them in the Appendix B.3.12.

Lemma 3.1. *Suppose (M) holds and $p_1(l|l) \geq p_2(l|l)$. Then, the all-pay contest has a unique equilibrium which has the following properties:*

- *the support of equilibrium strategies of each type is disjoint from but connected to the other type of the same player,*
- *the highest score played in equilibrium, $\Delta_{l,l}$, is in the strategy support of any l-type,*
- *the joint support of player 1's strategies is $(0, \Delta_{l,l}]$,*
- *the joint support of player 2's strategies is the same as that of player 1 plus an additional mass point at 0, in case $p_1(l|k) \neq p_2(l|k)$ for some k,*
- *both players play mixed strategies with piecewise constant densities on at most three subintervals of $(0, \Delta_{l,l}]$.*

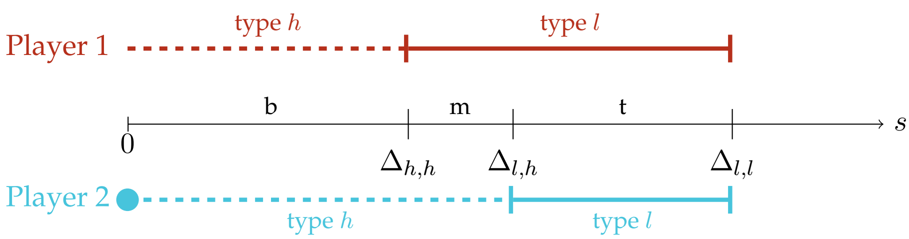

FIGURE 3.1: *Strategy support of player 1 and 2 with type-dependent priors.*

The Lemma is a direct application of Siegel (2014) to our setting. Figure 3.1 summarizes the equilibrium strategies. The horizontal axis depicts the score s. The dark-red and the light-blue line denote equilibrium strategy support for both players if player 1 is more likely to have low-cost. Player 1 (dark-red line at the top), type h (dashed part), is indifferent for all scores on the bottom interval b from 0 up to and including $\Delta_{h,h}$. This is the lower bound for the score of $1l$ (solid part) who is indifferent on all scores on intervals m and t up to and including $\Delta_{l,l}$ given the strategy of player 2. Player $2h$ (light-blue dashed line at the bottom) is indifferent between a score of 0 (indicated by the dot) and on intervals b and m up to and including $\Delta_{l,h}$. Player $2l$ is indifferent on interval t. If players become ex-ante symmetric, interval m vanishes, the mass point at 0 disappears, and strategies become fully symmetric.

There are no pure-strategy equilibria: whenever one player scores on a singleton only, it is either optimal to marginally overscore this value or to score 0 instead. There are several relevant properties of this mixed-strategy equilibrium. First, the highest score obtained by both players is the same. If one player was to strictly overscore her opponent, she could always deviate by reducing her score to the highest possible score of her opponent. Such a deviation does not reduce the probability of winning, but reduces the cost of the score.

Second, choices in all-pay contests are similar to strategic complements: whenever the likelihood of player $1l$ increases, player $2l$ reacts by scoring more aggressively. As l-types share the upper bound in their strategies, $2l$ has a higher average score than player $1l$.

Third, for every information structure at least one h-type player receives 0-utility in expectations. This player is always the ex-ante weakest player-type combination, here player $2h$. If this is not the case, no player would score exactly 0 with positive probability. But then, whatever the lower bound of the joint support, scoring at this lower bound yields a negative utility, which can always be avoided by deviating to a score of 0.

If player $2h$ has a mass point at 0, player $1h$ receives strictly positive utility as every score arbitrarily close to 0 guarantees her to win if player $2h$ decides to score 0.

Overall, the equilibrium actions in the all-pay contest depend on the belief about both the opponent's type, and the opponent's action, where the latter is a function of the opponent's beliefs. Thus, expected utilities depend on the entire belief structure. The following corollary to Lemma 3.1 defines the expected contest utilities in closed form.

Corollary 3.1. *Under the assumptions of Lemma 3.1, and $p_i(l|k) > 0$, the expected contest utilities are*

$$\begin{aligned} U_1(l) = U_2(l) = 1 - c_l\, \Delta_{l,l} > 0, \\ U_1(h) = p_2(h|h) F_{2,h}(0), \\ U_2(h) = 0. \end{aligned} \tag{U}$$

Moreover, utilities are linear in beliefs, if beliefs are type-independent. If beliefs are symmetric, $F_{2,h}(0) = 0$.

The utility of the low-cost types is a direct consequence of the common highest score. Both players win with probability 1 if they score at $\Delta_{l,l}$ and have cost $c_l \Delta_{l,l}$. On all other scores in their support they must be indifferent. The utility of the high-cost type of player 1 is derived as she always wins against those high-cost types that score 0 even if she scores arbitrarily close to 0. High-cost types of player 2 score 0 with probability $F_{2,h}(0)$ which gives them utility 0. If beliefs become type-independent, that is $p_i(l|l) = p_i(l|h)$, the upper bound, $\Delta_{l,l}$, and the mass on 0, $F_{2,h}(0)$, is linear in beliefs. If beliefs are symmetric between players, that is $p_1(l|k) = p_2(l|k)$, the mass point on 0, $F_{2,h}(0) = 0$ and $U_1(h) = 0$.

3.4.2 Deviator Payoffs in the Continuation Game

As players in our model differ only with respect to their cost in the contest, it is important for incentive compatibility to characterize the post-deviation continuation game. It needs to be assessed how players' actions and utilities change in case of breakdown conditional on a false report during the reporting stage. A false report introduces non-common knowledge of beliefs between the players. The deviating player knows about her deviation and assigns correct beliefs to her opponent. The non-deviating player and the mediator, on the other hand, are unaware of the deviation and incorrectly predict the deviator's beliefs. The wrong prediction affects actions, expected contest utilities, and thus incentive compatibility.[16]

Lemma 3.2. *Assume (M) and $p_1(l|l) \geq p_2(l|l) > 0$. All player-type combinations but player $1h$ are weakly better off in their respective deviation contest. Player $1h$ is strictly*

[16] The deviator of course correctly predicts the wrong prediction of the non-deviator, and so on.

worse off in the deviation contest if and only if the probability of facing a high-cost type in her deviation contest is strictly smaller than in her on-path contest.

Lemma 3.3. *Assume (M) and $p_1(l|l) \geq p_2(l|l) > 0$. Then, exactly one type of each player is strictly better off in the deviation contest than in the on-path contest if and only if the beliefs the player holds are* not *type-independent. If beliefs are type-independent, no player is better off in the deviation contest.*

Lemmas 3.2 and 3.3 state that the only situation in which no player-type prefers the deviation contest to the on-path contest is when beliefs, $p_i(l|m)$, are independent of the reported type m. To understand the intuition, let us first define the two types of contest.

Definition 3.1. *On-path contest:* the contest is called on-path contest if the belief structure is such that any player i, type k, holds belief $p_{-i}(l|k)$ about player $-i$. Further, the belief that each player and type holds is common knowledge.

Definition 3.2. *Deviation contest:* the contest is called deviation contest of player i k if player i, type k holds a belief $p_{-i}(l|\neg k)$ that is the same belief that player i, who is *not* k, holds on-path. This belief is called the deviator's belief. Player $-i$, however, holds her on-path belief $p_i(l|k)$ about player i. Thus, generically, there is no common knowledge of beliefs in this contest.

A direct consequence of non-common knowledge of beliefs is that the deviating player is no longer indifferent between several scores. The non-deviating player chooses her strategy to make an *on-path opponent* indifferent on some interval. The deviator, however, has a different belief about the non-deviator than the *on-path opponent* and is thus *not indifferent*. Decisions are similar to strategic complements, such that a too aggressive choice of the non-deviator leads the deviator to pick an aggressive response. If the choice is too soft, the deviator picks a soft response. The best response is generically a singleton.

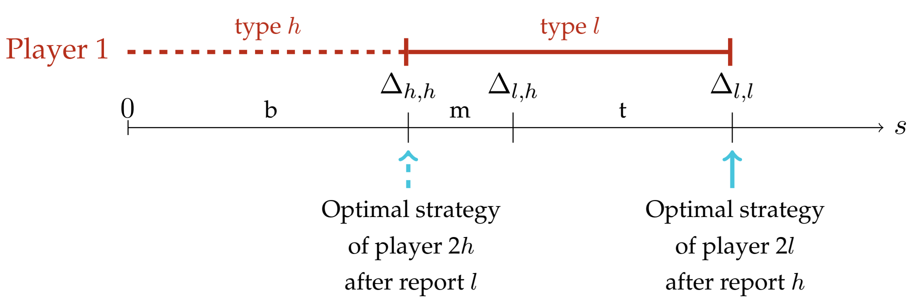

FIGURE 3.2: *Optimal behavior in the deviation contest of player 2 if $p_1(l|h) > p_1(l|l)$. Notice that the deviation strategies are conditional on $2l$ reporting h and $2h$ reporting l without player 1 noticing.*

Figure 3.2 illustrates the optimal strategies for player 2's deviation contest in case it is more likely that l-types appear after an h-report, i.e. $p_1(l|h) > p_1(l|l)$. The horizontal axis describes the scores, the dark-red line the strategy of player 1, which is the same as in equilibrium. The light-blue, dashed arrow points to the unique best response of player $2h$ who reported l, the solid arrow to that of player $2l$ who reported h.

If the probability that the opponent has low cost is larger in the deviation contest, the deviating l-type decides to score more aggressively. By the common upper bound in the strategy support, scoring above the highest score, $\Delta_{l,l}$, is never beneficial. Thus, her optimal strategy in the deviation contest is to score at $\Delta_{l,l}$ and to win with probability 1, if she is more likely to meet an l-type. Therefore, her utility is the same as on path, where she wins with probability 1 at a score $\Delta_{l,l}$ which is part of her equilibrium strategy.

Whenever reporting h increases the likelihood to meet an l-type opponent for player 2, reporting l must increase the likelihood to meet an h-type, i.e. $p_1(h|l) > p_1(h|h)$. Similar to the case of $2l$, a deviation by $2h$ makes her increase the score against an h-type (interval b in Figure 3.2), but decrease it against an l-type (interval m in Figure 3.2), since those occur less likely. Thus, her optimal response is $\Delta_{h,h}$ which leads to a win against all h-types. High-cost types occur with higher probability as $p_1(h|l) > p_1(h|h)$, and hence, $2h$ prefers the deviation contest to the on-path contest.

Low-cost players are never worse off in the deviation contest, as they can always score at the top. Moreover, player $2h$ is not worse off either as she can secure her on-path utility of 0. The only player that can be worse off in the deviation contest is player $1h$, if she expects to meet less $2h$. She then softens her bid to 0 and wins by the tiebreaker but suffers from the low probability of meeting $2h$.

Having discussed both on-path and post-deviation behavior in the continuation game, we shorten notation and use $U_i(k|m)$ to describe the expected utility that player i, type k enjoys in the contest stage if she reported to be type m and behaves optimally thereafter.

3.4.3 Rewriting the Problem

We now turn to the problem of the designer. Note that the problem is highly nonconvex and standard techniques do not apply. To be able to characterize the solution we need to transform it to a tractable problem. We do so in several steps. As the transformation is a series of technical issues we proceed as follows. First, we state the proposition describing the reformulated problem. Second, we state the original problem. Third, we provide a brief, non-technical comment on each transformation step in the main text. We refer the interested reader to Appendix B.1 for the corresponding detailed description of the transformation including the intermediate lemmas.

Proposition 3.1. *Any ex-post implementable, individually feasible and incentive compatible solution to*

$$\min_P Pr(\Gamma) = \min_P R(P)\gamma^*(P) \tag{P1'}$$

is also a solution to the mediator's problem if and only if $\gamma^*(P) \leq 1$, *where* $R(P) = Pr(\Gamma)/\gamma(l,l)$.

The proposition states that an equivalent formulation of the mediator's problem exists. In it, she optimizes over the set of breakdown beliefs, $P = \{p_1(l|l), p_2(l|l), p_1(l|h)\}$, instead of the set of shares and breakdown probabilities, $\mathcal{X} = (G, X_1, X_2)$. The remaining breakdown belief about player 2, $p_2(l|h)$, is implicitly defined by P and

Bayes' rule. The rewritten problem comes at the cost of two additional, technical constraints, namely ex-post implementability and individual feasibility. We are going to discuss these constraints below.

The Original Problem of the Mediator. As the mechanism needs to pass a ratification stage it is not necessarily without loss of generality to assume full participation. Given the payoff structure of the litigation game, however, we can use a result of Celik and Peters (2011) to conclude that full participation is indeed optimal in our setting, the corresponding lemma stating this result is included in Appendix B.1. Given full participation, the mediator's problem is

$$\min_{\mathcal{X}} Pr(\Gamma) = \min_{\mathcal{X}} (p, (1-p)) \cdot G \cdot \begin{pmatrix} p \\ (1-p), \end{pmatrix} \tag{$P1$}$$

subject to the following sets of constraints for all $i \in \{1, 2\}$ and $k, m \in \{l, h\}$

$$\Pi_i(k|k) \geq V_i(k), \tag{PC_i^k}$$

$$\Pi_i(k|k) \geq \Pi_i(k|m), \tag{IC_i^k}$$

$$x_1(k_1, k_2) + x_2(k_1, k_2) \leq 1, \quad x_i(k_1, k_2) \geq 0,$$

$$0 \leq \gamma(k_1, k_2) \leq 1,$$

where $\Pi_i(k|m)$ describes the expected total payoff of a participating player i, type k given she reports m. $V_i(k)$ describes the value of vetoing the mechanism for player i, type k. The first set of constraints are participation constraints, (PC_i^k), indicating that each player and type should prefer to participate in ADR over vetoing. The second set, the incentive compatibility constraints (IC_i^k), state that it is optimal for each agent to announce her true type. The third set of constraints prohibits additional payments by the agents or the mechanism and ensures a balanced budget. Finally, the last set of constraints ensures that breakdown probabilities are between 0 and 1.

Value of vetoing. To determine the outside option we need to define the continuation equilibrium of the litigation game after a veto by either of the parties in the ratification stage. High-cost types do not receive any payoff after a veto and are thus always at least indifferent to participate in ADR. Low-cost types' value of vetoing depends on the choice of beliefs after vetoing. In our case any choice of these off-path beliefs after vetoing which satisfy the intuitive criterion leads to the same value of vetoing: the expected litigation payoff under the prior p.[17]

Whenever the value of vetoing is smaller than $1/2$ for low-cost types, however, the mediator could offer parties a sharing rule of $(1/2, 1/2)$ for each type-realization and settle all cases. To make the problem interesting we make the following assumption.

Assumption 1. The low-cost types' value of vetoing is strictly above $1/2$.

Assumption 1 translates into the following condition on parameters: $\kappa > (2-2p)/(1-2p)$.

[17]This is a direct consequence of the low-cost types' contest utilities being a function of the weaker players' probability to have low-cost in case of type-independent beliefs. Any deviation belief satisfying the intuitive criterion, makes the non-deviating player the weaker one. Thus, the relevant belief remains constant at p.

Expected payoff. The expected payoff from participation, $\Pi_1(k|m)$, has two components: the expected value of successful settlement and the expected value of mediation breakdown and subsequent litigation. Thus,

$$\Pi_i(k|m) = z_i(m) + \gamma_i(m)U_i(k|m), \tag{3.1}$$

where message m leads to a value of settlement, $z_i(m)$, and a value of breakdown $\gamma_i(m)U_1(k|m)$. The expected contest probability, $\gamma_i(m)$, is a convex combination of the breakdown probabilities conditional on the opponents type

$$\gamma_1(m) = p\gamma(m,l) + (1-p)\gamma(m,h),$$

the value of settlement is a convex combination of realized shares and settlement probabilities

$$z_1(m) = p(1-\gamma(m,l))x_1(m,l) + (1-p)(1-\gamma(m,h))x_1(m,h),$$

and analogously for player 2. Equation (3.1) shows how optimal mediation relies on the litigation game. While the value of settlement, z_i, is similar to transfers in standard mechanism design, the utility of the contest continuation game is the screening device.

Step 1: Reduced-Form Problem à la Border (2007). In this step we make use of a procedure introduced by Border (2007) to reduce the problem from realized values to expected values. The reduced form problem has the advantage that the exact composition of the settlement shares, X_i, becomes irrelevant and we can use the settlement values, $z_i(\cdot)$, directly as choice variables. To ensure a feasible X_i, reducing the problem introduces two additional constraints: an individual feasibility constraint, (IF), and an ex-post implementability constraint, (EPI). The first constraint states that each player cannot get more than the whole good in case of settlement. The second constraint guarantees that the total amount of value distributed to a given type-profile does not exceed the total probability of any of the types within that profile occurring.

Step 2: Backing out Expected Settlement Shares. In the second step, we make use of the fact that we can assume without loss of generality that both the high-cost types' incentive compatibility constraints and the low-cost types' participation constraints are binding. The latter follows naturally from the values of vetoing, that is the fact that low-cost types need to be compensated to take part in ADR. Binding incentive compatibility for high-cost types follows from their low expected payoff in litigation: it provides an incentive to mimic low-cost types to get their settlement value. The binding constraints allow us to eliminate all settlement values, as they can be expressed in terms of breakdown valuations.

Step 3: From Breakdown Probabilities to Breakdown Beliefs. This step uses that breakdown beliefs are homogeneous of degree 0 with respect to the set of breakdown probabilities, G by Bayes' rule. Thus, the set of breakdown *beliefs* defines the set of breakdown *probabilities* up to a constant. We choose this constant to be $\gamma(l,l)$ such that all other breakdown probabilities are defined relative to $\gamma(l,l)$. This allows us to eliminate all breakdown *probabilities* but $\gamma(l,l)$, and replace them by breakdown *beliefs*.

Step 4: Eliminate $\gamma(l,l)$ via expected feasibility. The final step is to eliminate

$\gamma(l,l)$. We use the fact an ex-ante feasible settlement rule is a necessary condition for individual feasibility, (IF). All expected breakdown probabilities increase linearly in $\gamma(l,l)$ by Step 3. Therefore, the mediator wants to set $\gamma(l,l)$ as low as possible, as long as the problem remains feasible in expectation. This introduces an equality constraints $\gamma(l,l) = \gamma^*(P)$ by which we replace $\gamma(l,l)$. The additional constraint $\gamma^* \leq 1$ ensures that $\gamma(l,l)$ remains a probability. This concludes the rewriting of the problem.

3.4.4 Optimal ADR-Mechanism

Having established the reduced problem (P1'), which is a problem of three choice variables only, we can now state the main result:

Theorem 3.1. *Suppose Assumption 1 holds. Then, any optimal mediation protocol has the following properties:*

- *on-path breakdown beliefs are type-independent, that is for any i it holds that $p_i(l|l) = p_i(l|h) =: \rho_i$,*
- *on-path breakdown beliefs are asymmetric, that is $\rho_i \neq \rho_{-i}$,*
- *both player's on-path breakdown belief is weakly larger than the prior, that is $\rho_i \geq p \; \forall i$,*
- *all type profiles $\{k_1, k_2\}$ have a breakdown probability that is strictly positive.*

Theorem 3.1 states that, independent of the primitives, any optimal protocol induces an information structure that is report-independent. In addition, although parties start perfectly symmetric, the mediation protocol should always be set up asymmetrically. At the same time the ADR protocol ensures that both parties appear to be at least as strong after mediation breakdown as they appeared before mediation. Therefore, the fraction of low-cost types is at least as high in a post-mediation contest as before the start of the game. Finally, the mediator needs to ensure that in principle any type profile can lead to a breakdown of mediation to get the above mentioned features.

To build intuition we organize the remainder of the section as follows. We first discuss the optimal solution to (P1') ignoring (IC_i^l) and $\gamma^*(P)$. We then reintroduce (IC_i^l) and later $\gamma^*(P) \leq 1$. Finally, we verify that the solution is implementable in the sense of Border (2007).

Recall that the assumption of player 1 appearing weakly stronger in the contest implies the following expected litigation utilities of the high-types: $U_2(h|h) = 0$ and $U_1(h|h) \geq 0$ with strict inequality whenever player 1 appears strictly stronger. Further, litigation utilities, $U_i(k|m)$, depend on breakdown *beliefs* and all expected breakdown *probabilities*, $\gamma_i(m)$, are linear in $\gamma(l,l)$. In addition, the following technical lemma is useful to keep in mind. It states that whenever it is more likely for player 2 to meet $1l$ after a report of l, the same is true for player 1 and vice versa.

Lemma 3.4. $p_1(l|l) > p_1(l|h) \Leftrightarrow p_2(l|l) > p_2(l|h)$ *if* $p_i(l|m) \in (0,1)$.

Part 1: Neglecting (IC_i^l) and $\gamma^*(P) \leq 1$. First, we want to argue that beliefs are type-independent. The basic idea is straight-forward: if the mechanism does not allow parties to influence the opponent's type distribution in case of breakdown, then there is no incentive for a false report. Similar to a second price auction, where expected payments are independent of the type report, the mediator ensures that the type distribution the player faces, and by that her contest utility, is independent of her type report.

Proposition 3.1 states a problem with the three breakdown beliefs, $p_1(l|l), p_2(l|l), p_1(l|h)$ as choice variables. Given Lemma 3.4 we can fix $p_2(l|l)$ and $p_1(l|h)$ for the upcoming argument and concentrate on $p_1(l|l)$ without loss of generality.

As the mediator cannot achieve full settlement by the participation constraint of the low-cost types and the high-cost types' desire to mimic them, she needs to strategically fail mediation to screen types. High-cost types need to be present in the contest to guarantee some utility for the low-cost player and to match her participation constraint. However, the high-cost players should have an incentive to avoid the contest to report truthfully. Thus, the probability of a high-cost player meeting another high-cost player after mediation breakdown, $p_i(h|h)$, should be smaller than the ex-ante probability of a high-cost type, $1-p$. Without a belief dependent outside option this effect typically drives $p_i(l|h)$ to 1 as in e.g. Hörner, Morelli, and Squintani (2015).

There is, however, a second, non-standard effect, changing utilities after breakdown. If breakdown is informative, i.e. $p_i(l|l) \neq p_i(l|h)$, the expected utility in the deviation contest might differ from the expected utility in the on-path contest.

Recall from Lemmas 3.2 and 3.3 that $U_i(h|l) > U_i(h|h)$ whenever it is more likely to meet a low-cost type under truth-telling than under deviation, that is whenever

$$p_i(h|h) < p_i(h|l) \Leftrightarrow p_1(l|l) < p_1(l|h),$$

due to the information advantage effect in the contest. This advantage vanishes as $p_1(l|l) \to p_1(l|h)$. If $p_1(l|l)$ increases further, player 2 receives no utility in the contest and therefore also no marginal breakdown utility from lying. Player 1, on the other hand, actually starts gaining utility again, as an intimidation effect becomes dominant. Player 1 appears to be much stronger in expectation than player 2. Thus, player 2 invests less into the contest which increases player 1's utility. Therefore, both deviation utilities have a minimum at type-independent beliefs.

Deviation utilities have a kink at type-independent beliefs by the all-pay contest assumption. The kink is a direct consequence of Lemma 3.3 as deviating high-cost players are only indifferent for type-independent beliefs. High-cost types score at the upper end of their on-path equilibrium strategy set for lower values of $p_1(l|l)$ and at the lower end for higher values of $p_1(l|l)$. Hence, for type-independent beliefs their utilities are non-differentiable and obtain a minimum. The left panel of Figure 3.3 plots the deviation utilities as a function of $p_1(l|l)$.

If we combine the effects on breakdown probabilities $\gamma_i(m)$ and contest utilities, we find that the minimum at type-independent beliefs prevails. The result can best be seen if we consider the marginal breakdown-value of lying. This breakdown value is the right-hand side of the following representation of the high-types incentive constraint,(IC_i^h),

$$z_i(h) - z_i(l) = \gamma_i(l)U_i(h|l) - \gamma_i(h)U_i(h|h).$$

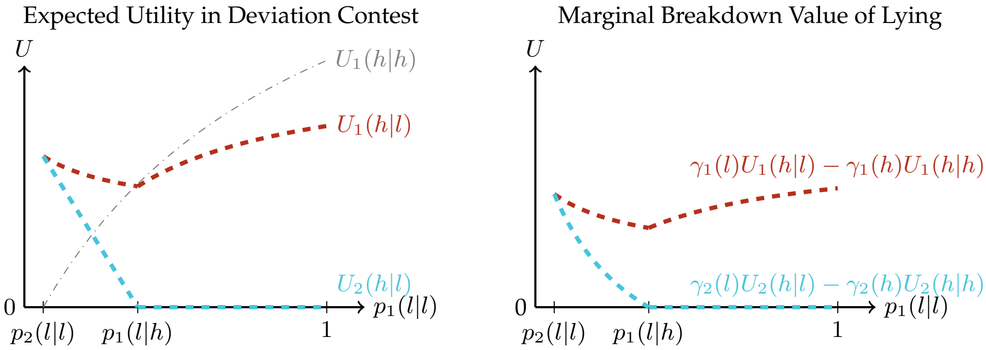

FIGURE 3.3: *The left panel depicts the high-types deviation utilities as a function of $p_1(l|l)$. The right panel depicts the marginal breakdown-value of lying. Red is for player 1, blue player 3. The gray line in the right panel is the on-path utility of the high-cost type of player 1.*

The left hand side can be interpreted as the marginal settlement value of truth-telling which matches the right hand side being the marginal breakdown value of lying. The right panel of Figure 3.3 displays the marginal breakdown value of lying and illustrates how the minimum property prevails and type-independent beliefs are optimal. We can thus simplify notation and define ρ_i to be the probability that player i is the low-type post-mediation.

Having established that beliefs are type-independent we can simplify the analysis using a corollary to the derivation of the breakdown beliefs.[18]

Corollary 3.2. *If beliefs are type independent, breakdown probabilities can be simplified to*

$$\gamma_i(l) = \frac{p}{\rho_{-i}}\gamma(l,l), \qquad \gamma_i(h) = \frac{(1-\rho_i)}{(1-p)}\frac{p}{\rho_i}\gamma_i(l), \qquad Pr(\Gamma) = \frac{p^2}{\rho_1\rho_2}\gamma(l,l).$$

Moreover, Corollary 3.1 allows us to write contest utilities with type-independent beliefs as

$$U_i(l|m) = (1-\rho_2)\frac{\kappa-1}{\kappa}, \qquad U_1(h|m) = (\rho_1-\rho_2)\frac{\kappa-1}{\kappa}. \tag{3.2}$$

These expressions are useful in the argument for asymmetry of the optimal mechanism which we turn to next. We discuss the general argument non-formally to provide a good understanding of the qualitative results. A more detailed and formal analysis is in Appendix B.2.

The main argument for asymmetry lies in the structure of a contest. A symmetric contest is expected to be tight: parties expect to be matched with an opponent of similar strength and the marginal value of investment is high. By contrast, an asymmetric contest appears to be less tight, and the marginal value of investment is lower for both parties. This imposes an externality, especially for the high-cost type of the ex-ante stronger player. Her opponent's high-cost type is going to increase her investment but remains at a utility of 0 as she is the weakest of all player-types. Thus, the stronger player's h-type can reduce the investment and still has a reasonable chance to win the contest as the opponent believes she likely faces a low-cost type. This effect can be seen by inspecting equations (3.2). If we start in a

[18] To be precise, Corollary 3.2 is a corollary to Lemma B.4 which is stated in Appendix B.1.

symmetric setting and unilaterally increase the belief put on player 1, then l-types would not benefit in terms of expected utilities and neither would $2h$. However, player $1h$ actually achieves a positive utility in such a case which she would not under symmetry.

Although only concerned about the probability of contest, the optimal ADR-mechanism uses this property of the underlying game to increase the breakdown utility of one of the high-cost types. This allows the mediator to reduce the settlement value that needs to be paid to this player which in turn increases the available resource for settlement. There is, however, a second effect that limits the extent to which the mediator can use this feature: as breakdown probabilities are derived in their *relative relation* to $\gamma(l,l)$ in problem $(P1')$, an increase in ρ_1 is effectively a decrease of the breakdown probability of high-cost types of player 1, $\gamma(h,l)$ and $\gamma(h,h)$. This implies, in turn, a decrease in the breakdown probability for *player 2l*, $\gamma_2(l)$, according to Corollary 3.2. While such a decrease has a positive effect on the objective, $Pr(\Gamma)$, it also leads to a decrease in player 2's breakdown utility. Thus, the mediator would need to increase player 2's settlement utility. Making the contest less resource intensive is therefore only optimal up to a certain point. This point balances the additional resources needed to finance the loss for player $2l$ and the gain from making the contest less resource-intensive. A similar argument is true for the other player-types.

To see the aggregate effect consider the expected settlement share paid to player i, z_i. The expected settlement share is a convex combination of the settlement share paid to the l-type to ensure participation and the settlement share paid to the h-type to ensure incentive compatibility. The shares are given by

$$z_2 = V(l) - \frac{1-\rho_2}{\rho_1}\frac{\kappa-1}{\kappa}p\gamma(l,l)$$

$$z_1 = \underbrace{V(l) - \frac{1-\rho_1}{\rho_2}\frac{\kappa-1}{\kappa}p\gamma(l,l)}_{\text{symmetric part}} + \underbrace{(\frac{p}{\rho_1} - \frac{p}{\rho_2})\frac{\kappa-1}{\kappa}p\gamma(l,l)\kappa}_{\text{asymmetric part}}.$$

The first part in z_1 is present in the symmetric case, too, while the second vanishes. Without the second part z_1 would be the anti-symmetric version of z_2 which would lead to endogenous symmetry. However, the second part provides a clear incentive for asymmetry driven by $U_1(h|h)$.[19] An increase in ρ_2 requires more resources to compensate the players than an increase in ρ_1. Thus, the optimal choice involves $\rho_1 > \rho_2$, that is player 1 appears relatively stronger in the contest. Finally, notice that the asymmetric part is always negative and thus, some asymmetry always saves resources. The next lemma states the findings up to this point.

Lemma 3.5. *Ignoring* (IC_i^l), (IF), (EPI) *and* $\gamma(l,l) \le 1$, *and assuming that* $\rho_1 \ge \rho_2$, *the unconstrained optimum of (P1') is achieved at*

$$\rho_1^* = \frac{1+p}{2} \qquad\qquad \rho_2^* = \frac{1-p}{2}.$$

Moreover, the optimal breakdown belief ρ_i^* *is independent of the opponents breakdown belief* ρ_{-i}.

[19] Notice that this part can also be written as $-Pr(\Gamma)U(h|h)$.

Part 2: Reintroducing (IC_i^l). Next, we reintroduce the low-cost type's incentive compatibility constraint, (IC_i^l). For type-independent beliefs and with (IC_i^h) satisfied this boils down to

$$(\gamma_i(l) - \gamma_i(h))U_i(h|h) \leq (\gamma_i(l) - \gamma_i(h))U(l|l). \tag{3.3}$$

A sufficient condition for this to hold is $\gamma_i(l) \geq \gamma_i(h)$, as $U(l|l) \geq U_i(h|h)$ by construction. For player 2 it is also necessary since $U_2(h|h) = 0$. Using Corollary 3.2, $\gamma_i(l) \geq \gamma_i(h)$ is equivalent to $\rho_2 \geq p$. Intuitively the reasoning is straightforward: suppose $\rho_2 \leq p$. The likelihood of breakdown must be larger when reporting to be an h type. By (IC_2^h), the value of settlement, $z_2(l) = z_2(h)$, is independent of the report and the low-cost type prefers to be sent to contest more often and would misreport. Thus, incentive compatibility requires $\rho_2 \geq p$.

Taking into account the results from Lemma 3.5, this means that (IC_i^l) is violated whenever $(1-p)/2 < p$ which holds if and only if $p > 1/3$. Note further that $\rho_1^* > p$ for all p and thus, (IC_1^l) never binds. As the optimal ρ_i does not depend on ρ_{-i}, we get the following lemma.

Lemma 3.6. *Ignoring (EPI) and $\gamma^*(P) \leq 1$, and assuming that $\rho_1 \geq \rho_2$, (IC_i^l) binds for player 2 if and only if $p \geq 1/3$. In this case the constrained optimum is achieved at*

- $\rho_1^* = \frac{1+p}{2}$
- $\rho_2^* = p$.

Lemma 3.6 states that the probability of breakdown for low-types is larger than the probability of breakdown for high-types, i.e. $\gamma_i(l) \geq \gamma_i(h)$. In such a case one individual feasibility, (EPI), which is one of the two constraints coming from the reduced form, is always satisfied. Appendix B.3.4 provides details on this.

Part 3: Full model. So far we have ignored that the scaling parameter γ^* is in fact always equal to the probability of breakdown for two low-cost types, $\gamma(l,l)$, in the original problem. Thus, we need to ensure that $\gamma^* \in [0,1]$ to guarantee that $\gamma(l,l)$ remains a probability.

Whenever the constraint $\gamma^*(P)$ binds, (IC_i^l) must hold, too. To see this recall

$$\gamma_i(l) = \frac{p}{\rho_{-i}}\gamma(l,l).$$

To ensure $\gamma_i(l) \in [0,1]$ even if $\gamma(l,l) = 1$ we need $p \leq \rho_{-i}$. Such a high post-breakdown belief ensures incentive compatibility by Lemma 3.6. If the ex-ante probability of low-cost types is high enough for (IC_i^l) to bind, the scaling parameter $\gamma^*(P) < 1$. Thus, $\gamma^* \leq 1$ does not change the results of Lemma 3.6. Next, recall that

$$\gamma^*(P) = \frac{\nu}{Q(P) - R(P)},$$

such that γ^* is increasing in ν for any P. The value of ν, in turn, is large for small p and large κ. Therefore, the solution computed in Lemma 3.5 violates $\gamma^* \leq 1$ if cost difference between low-cost and high-cost types are high, or the probability to have high-cost is small.

To compensate this, the mediator can decrease either ρ_i. As in the discussion of Lemma 3.5 such an operation increases the resources available for distribution in settlements and allows to reduce γ^*.

Given small values of the prior, p, the optimal breakdown belief ρ_i *without* considering the γ^*-constraint is strictly larger than p, and thus the mediator reduces both beliefs, ρ_1 and ρ_2, simultaneously up to the point at which one equals the prior, i.e. $\rho_2 = p$. If this does not suffice to make $\gamma(l,l)$ feasible, the mediator decreases the belief on player 1, ρ_1, further until $\gamma^*(P) = 1$. It turns out that the remaining Border-constraint, (EPI), holds at any such point and ex-post implementation is thus possible. Combining all results allows us to make a statement about any set of parameters, κ and p. The characterization is given in the next lemma which concludes the argument for Theorem 3.1.

Lemma 3.7. *Consider without loss of generality only $\rho_1 \geq \rho_2$. Fix some κ such that Assumption 1 hold and assume that the mediator cannot release public signals. Then there are three cutoff values p', p'' and p''' such that the optimum of the minimization problem is either 0 or satisfies*

- (IC_2^l) *and therefore $\rho_2 = p$ with equality only if $p \notin (p', p''')$,*
- $\gamma(l,l) \leq 1$ *with equality only if $p \leq p''$,*
- $2p < \rho_1 \leq (1+p)/2$ *where the last holds with equality only if $p \geq p''$.*

The cutoffs are given by:

$$
\begin{aligned}
p' &= \frac{1}{6(\kappa-1)}\left(\kappa - 8 + \sqrt{28 - 4\kappa + \kappa^2}\right), \\
p'' &= \frac{1}{2+3\kappa}\left(2(\kappa-1) - \sqrt{8 - 4\kappa + \kappa^2}\right), \\
p''' &= \frac{1}{3}.
\end{aligned}
$$

The cutoffs describe the main characteristic of the optimum. For low p the mediator offers low-cost types a litigation utility post breakdown which is smaller than their value of vetoing, i.e. $\rho_2 > p$. To do this l-types need a high enough settlement share which the mediator finances by reducing the overall breakdown probability by increasing γ^*. However, for very low p not even $\gamma^* = 1$ suffices as $V(l)$ is too high. To account for the constraint, the mediator decreases both breakdown probabilities, ρ_2 and ρ_1. However, ρ_2 cannot fall below p as this would violate both (IC_2^l) and $\gamma_i(k) \leq 1$. Thus, for very low p, the mediator chooses $\rho_2 = p$ and adjusts ρ_1 accordingly.

As the prior p increases, the solution ρ_2 increases, too, and $\rho_2 \geq p$ does not bind anymore. The resource constraint, $\gamma^* \leq 1$, however, still does. If p is larger than p'', the solution of Lemma 3.5 can be implemented directly. For $p > 1/3$, on the other hand, low-cost types of player 2 have an incentive to misreport given the protocol from Lemma 3.5 which means that (IC_2^l) binds and the belief on player 2 is set to the prior, $\rho_2 = p$. The left panel of Figure 3.4 illustrates the findings. The dashed line plots the optimal protocol according to Lemma 3.5 whereas the solid line is the full model.

3.5 Discussion of the Results

Comparative Statics. Figure 3.4 depicts the probability of litigation under the optimal mechanism both as a function of the prior, p (left panel), and as a function of the

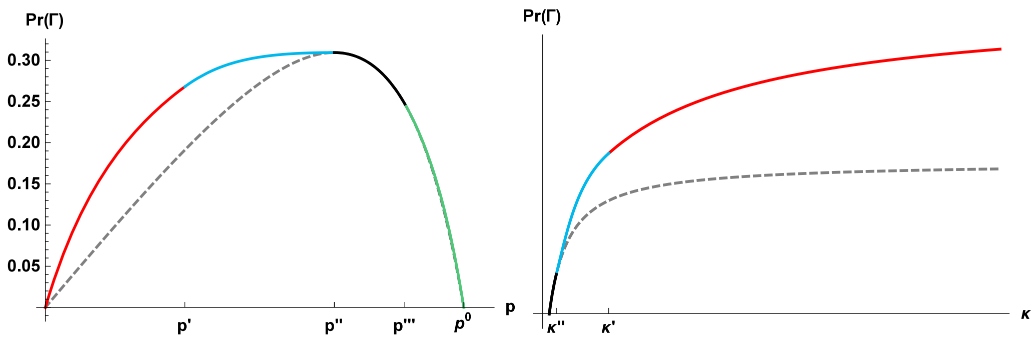

FIGURE 3.4: *Ex-ante probability of the contest as a function of p (left panel) and κ (right panel). The dashed line describes the situation of the unconstraint problem ($P1'$) as in Lemma 3.5. The green solid line corresponds to Lemma 3.6. All solid lines together display the result of Lemma 3.7.*

distance between low and high cost, κ (right panel). The different colors indicate the different regimes as discussed in Lemma 3.7. Red and blue (for $p < p''$) denote the areas in which the resource constraint, $\gamma^* \leq 1$, binds; green (to the right of p''') is the area in which $2l$'s incentive constraint binds and black is the area in which ($P1'$) is solved "unconditionally" as in Lemma 3.5. p^0 indicates the point at which Assumption 1 starts to fail and the mediator achieves full settlement for $p > p^0$. For comparison, the dotted line depicts the solution ignoring (IC_i^l) and $\gamma^* \leq 1$.

As expected, the probability of litigation increases in the distance between high-costs and low-costs. As the low-cost type's cost advantage increases, it becomes more expensive to compensate her for participation and thus the mediator initiates breakdown more often. The relationship with respect to the prior is non-monotone. When chances to meet a low-cost type are small, litigation can effectively be avoided. Although low cost types require a large compensation for a settlement, the mediator can grant this as she needs to pay this compensation seldom. As the ex-ante probability of low-cost types increases the mediator must pay the compensation more often, but at the same time the amount decreases. The result is an inverse U-shaped relationship between the prior and the probability of litigation.

In addition, comparative statics show that ADR is a very effective tool. In our setup the mediator can settle the majority of the cases for any set of parameters, p and κ. The next proposition summarizes these findings.

Proposition 3.2. *Under the optimal mediation protocol, the ex-ante probability of breakdown is never greater than $1/2$. Moreover, the probability of breakdown is increasing and concave in κ while it takes the form of an inverse U-shape in p.*

Next, we want to discuss how the asymmetry translates to the different outcome variables. A first result is straightforward and a direct consequence of Theorem 3.1: low cost types experience breakdown more often than high-cost types. Moreover, player $1l$ is sent to court more often than player $2l$ as the belief on player 1 is larger than on player 2. Since the participation constraint binds, both low-cost type players experience the same utility in expectations. However, the contest utility is the same for both low-cost types and smaller than the value of vetoing, $V(l)$, as low-cost types are more likely after breakdown than in the initial population. Thus,

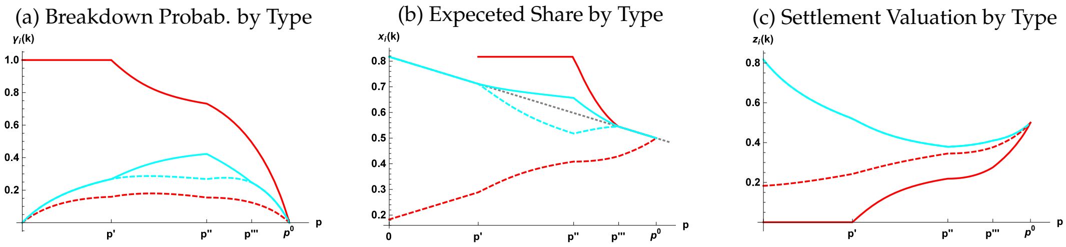

FIGURE 3.5: (a) *Expected Contest Probability,* (b) *Expected Share conditional on settlement taking place, and* (c) *Valuation of Settlement by player-type as a function of the ex-ante probability of being a low-cost type. Solid lines depict low-cost types, dashed lines depict high-cost types. Dark-red is player 1 and light-blue is player 2. The dotted gray line in* (b) *is the value of vetoing for low-cost type players. In* (c)*, player 2h has the same settlement value as 2l by incentive compatibility.*

player $2l$, who is sent to court less often, receives a smaller expected share than player $1l$. For high-cost types the intuition is the other way around. Player $2h$, who experiences no utility in contest post-mediation, is compensated with a larger amount than $1h$. The next proposition states that this is the case for all parameter values. Thus, player 1, who is stronger in the contest, expects a less favorable settlement contract than player 2 who, in turn, faces a more difficult task to win the litigation process after breakdown.

Proposition 3.3. *Both the pre-mediation probability of being sent to court during mediation and the expected share conditional on settlement are largest for player $1l$ and smallest for player $1h$.*

Figure 3.5 illustrates the results of Proposition 3.3 as a function of the prior distribution. The left panel (a) describes breakdown utilities, the middle panel (b) expected shares conditional on settlement, $x_i(m) \equiv z_i(m)/(1 - \gamma_i(m))$, and the right panel (c) the settlement valuation. Dark-red lines are for player 1 and light-blue lines for player 2. Dashed lines indicate high-cost types, solid lines indicate low-cost types. The linear, gray, dotted line in panel (b) denotes the value of vetoing for the l-type, $V(l)$.

If the probability of low-cost types is very small, the mediator sends one of the two low-cost types to litigation with certainty to ensure that the resource constraint holds. As the probability of low-cost players increases, the pressure from the resource constraint relaxes as compensation for low-cost types declines. The mediator thus wishes to implement a less asymmetric solution. As p increases further, the mediator can in fact reduce the probability of litigation for all types up to the point where Assumption 1 seizes to hold and the problem therefore becomes trivial.

Another feature of our model is that we are able to evaluate the consequences of the mediation decision on the litigation process. As litigation in our model is a strategic game with actions that depend both on first and second order beliefs, we should not expect players to play the same strategies as in litigation without a preceding mediation stage. Indeed the mediation attempt changes the belief structure of the opposing parties in two ways: (1) it increases the likelihood for both players to meet a low-cost type in court and (2) it introduces an asymmetry that makes player 1 more likely to be the low-cost type than player 2.

The first effect clearly makes competition more intense as litigants are afraid that the opponent can and will produce good evidence. The second effect works in the other direction, since the high-cost type of player 2 has little chance of winning in court. She refuses to compete at all from time to time and gives away the good for free. The second effect exceeds the first, if player 2's likelihood of being the low type is the same as the prior, that is if $\rho_2 = p$ or $p \neq (p', p''')$. In such a case the incentive to downsize investment in evidence due to the asymmetry between players always supersedes the incentive to increase investment in evidence due to the higher probability of low-cost types and we would see lower legal expenditure post-mediation.

Proposition 3.4. *Assume parameters are such that $p < p'$ or $p > p'''$. Then, the sum of expected legal expenditures after breakdown never exceeds the sum of expected legal expenditures if mediation did not exist.*

Outside this range no clear statement can be made other than that for any κ there exists a possibly empty interval $(\hat{p}, \check{p})$, with $\hat{p} \geq p'$ and $\check{p} \leq p'''$. Only in this interval, the expected legal expenditure after breakdown is higher than without mediation.

3.6 Extensions

Pre-trial Bargaining. The traditional law and economics literature focuses mainly on bilateral settlement negotiations. Typically, these bargains are modeled as a simple take-it-or-leave it bargaining game (Posner, 1996; Shavell, 1995; Schweizer, 1989). For illustration assume the following bargaining procedure close to Schweizer (1989): one player (Sender) makes a take-it-or-leave-it offer to the other player (Receiver) who decides whether to accept or reject the offer. Upon rejection both players update their beliefs and proceed to litigation.

To compare our results, notice first that by the revelation principle and Lemma B.1, the equilibrium rejection channel is absent. Pre-trial negotiations thus cannot outperform the result of the mechanism.

As in the mediation mechanism, off-path beliefs play a crucial role in the bargaining game. The actions in the contest are based on the belief structure as discussed above.

The solution concept of perfect Bayesian Nash equilibrium allows to freely choose beliefs put on the deviator at the first node of deviation, but requires Bayes' rule thereafter. Any bargaining equilibrium that performs as well as the mediation mechanism replicates outcome utilities of the mechanism and is furthermore equipped with a set of off-path beliefs that deter any deviation by any player. It turns out that no off-path belief exists such that the bargaining can replicate the mediator's solution as long as Assumption 1 holds.

Proposition 3.5. *Independent of the off-path belief structure, take-it-or-leave-it bargaining leads to a strictly higher probability of litigation than the optimal mediation mechanism provided that Assumption 1 holds.*

The intuition behind the result is that a low-cost Sender could always profitably deviate by proposing an arbitrarily small share ϵ to Receiver. Then, given any belief Receiver holds after observing this deviation, she either accepts the share which gives Sender a higher utility than in the optimal mechanism, or rejects the share if she thinks Sender is weak. Assuming a weak Sender, however, induces her to score softer than in the litigation game under priors. By strategic complementarity, Sender scores softer as well. But then, Sender expects a higher utility as winning is less costly. Thus, it is not optimal for a low-cost Sender to reproduce the outcome of the optimal mechanism: the incentive to deviate from the mechanism leads to a higher breakdown probability in expectations.

This shows the importance of a third-party who manages the information flow. With direct bargaining, Receiver always interprets Sender's proposal as a signal and Sender cannot commit to abstain from signaling via her proposal. A neutral third-party can overcome this adverse selection problem and thus improves upon bilateral negotiations.

Asymmetric Players. Asymmetric players do not change any of the results obtained. The reason is that the mediator would always treat the ex-ante stronger player as "player 2", i.e. the player that gets the better settlement conditions. The ex-ante weaker player accepts a small settlement share, since she fears a strong opponent in litigation. The weaker player, however, is compensated for the small share with a favorable contest after breakdown. Thus, while the ex-ante weaker player is strong post-breakdown, she agrees to settlement-contracts that favor her opponent. With such a protocol the mediator is still able to solve the majority of the cases. A key result of our analysis is, however, that we get asymmetric results even with symmetric players.

Different forms of commitment. So far we have assumed that both players can fully commit to the proposed mediation protocol. In particular, once the mechanism is accepted, parties commit to only go back to litigation if the mediator tells them. In reality this is not always the case. Many jurisdictions demand that parties can unilaterally opt-out of ADR at any point to return to litigation. We discuss two stages at which parties can unilaterally decide to break down ADR. The first is a situation in which they can leave after the mechanism has told players' their *expected* share conditional on settlement. We call this commitment structure post-ADR individual rationality (PAIR). The second commitment structure is that parties can veto the mechanism after they have learned their *realized* share conditional on settlement. We call this ex-post individual rationality (EPIR).

The mediation protocol developed in Section 3.4 does not directly carry over to PAIR and EPIR. In fact, given these commitment schemes, the mediator profits from the ability to communicate to parties even after ADR breaks down. If this is the case, the mediator can give parties non-binding recommendations for the play of the contest and by that restore the outcome under full-commitment. The modified game thus follows a slightly enhanced timeline:

1. the mediator commits to $\mathcal{X}$ and recommendation structure Σ; players learn their types,
2. players send a message m_i to the mediator,
3. the mediator *privately* announces a share x_i according to $\mathcal{X}$ to each player i,
4. players accept/reject the share,
5. players receive a recommendation σ_i by the mediator,
6. if either of the players rejected her offer, the contest is played under updated beliefs.

Note that since the mediator first observes the behavior of the players with respect to the announced share she has the ability to detect a deviation in this stage (other than in the reporting stage). To restore the result of Section 3.4 the mediator uses the following slightly more sophisticated mechanism(s).

To find the optimal PAIR mechanism, we need to define a convex combination of the protocol derived in Section 3.4 and its mirror image switching roles of player 1 and 2. Define $\hat{\mathcal{X}}_\lambda$, a mediation protocol such that $\mathcal{X}_i$ applies with probability λ and $\mathcal{X}_{-i}$ with probability $(1-\lambda)$. $\mathcal{X}_i$ denotes a mediation protocol similar to the one discussed in Theorem 3.1. When mediation is successful, player i is treated as "player 1". To trigger litigation in this protocol, the mediator offers a share of 0 to at least one of the players. This share is going to be rejected such that parties move to the litigation game.

To ensure EPIR we need that, in addition, the mediator sends both parties to contest irrespective of their reports with probability $\epsilon > 0$. Thus, we define $\hat{\mathcal{X}}^\epsilon_\lambda$ to be a mediation protocol such that with probability ϵ players are send to court and with probability $(1-\epsilon)$ the mediator executes $\hat{\mathcal{X}}_\lambda$. This is sufficient to ensure the following two results.

Proposition 3.6. *There exists a signal Σ, such that an incentive compatible PAIR mechanism $(\hat{\mathcal{X}}_{1/2}, \Sigma)$ has the same breakdown probability $Pr(\Gamma)$ as the mechanism $\mathcal{X}$ under interim individual rationality.*

Proposition 3.7. *For any $\delta > 0$, there exists a signal Σ and an $\epsilon > 0$, such that an incentive compatible EPIR mechanism $(\hat{\mathcal{X}}^\epsilon_{1/2}, \Sigma)$ achieves a breakdown probability $Pr(\Gamma)^\epsilon < Pr(\Gamma) + \delta$, where $Pr(\Gamma)$ is the optimal breakdown probability of the mechanism $\mathcal{X}$ under interim individual rationality.*

To gain intuition observe the following. First, with both PAIR and EPIR the mediator can trigger the play of a contest by offering at least one party an unacceptable share as rejection leads to contest. Second, the mediator achieves the result by obfuscating two issues: the role of the player and the relevance of her decision.

The latter derives from the possibility that the mediator wants to trigger contest play and has offered the player's opponent an unacceptable share. As both do not know which litigant takes the role of player 1, and who is offered the trigger share 0, she cannot learn much from her own offer. As the conditional distribution post-breakdown is on-path revealed via the signal σ, obfuscation is only payoff relevant in deviation games. Deviation is, however, only detected by the mediator, *not* by the non-deviator. Thus, the mediator can react to deviation by sending the deviator a signal of a strong non-deviator to punish her. This suffices to get the same result as under full-commitment.

In the case of EPIR the mediator is more constrained as revealing the ex-post share $x_i(k_1, k_2)$ to player i allows for more inference by the player. For some parameter values it might be the case that certain constellations do not settle on-path. Thus, the mediator might have a degenerate belief after some proposed realized shares which makes the procedure of PAIR impossible. The mediator can use another option instead, though. She can commit to initiate breakdown for any type-profile with a small probability ϵ and to send a fully informative signal thereafter. In such a case both parties can end up with 0 expected utility after breakdown. If the mediator commits to signal this event to the non-deviator after any deviation, the non-deviator will always invest an amount large enough to effectively punish the deviator. As $\epsilon \rightarrow 0$ the mechanism converges to $\hat{\mathcal{X}}_\lambda$ and the resulting probability of breakdown is arbitrarily close to that of the mechanism described in Theorem 3.1.

Nonetheless, allowing the types to go back to court after all uncertainty has unraveled would naturally lead to a different result. Typically however, once a detailed settlement agreement has been signed by both parties, it is hard to imagine a legal system that allows parties to overturn this contract simply because they have learned that they might have a good chance to beat the opponent.

3.7 Conclusion

In this paper we characterize optimal Alternative Dispute Resolution (ADR) in the shadow of the court. We show that optimal ADR is always asymmetric and offers one player an advantage after breakdown and the other one an advantage under settlement. We show that the optimal information structure post-ADR is completely independent of the players' report, but conditions only on their identity. Such a mechanism prevents players from misreporting to achieve an informational advantage.

We find that a litigation-minimizing ADR-protocol is highly effective and solves the majority of cases. The effectiveness indicates that mandatory ADR should be considered by all courts to reduce the prevalent stress on judges and court's backlog of cases. In addition, the asymmetry of the optimal mechanism implies that regulators should act carefully when defining their notion of fairness for mediation protocols. The same holds true for discretionary policies: mediators should always have the possibility to talk to the disputants in private as this eliminates commitment problems on the disputants side. Finally, we show that mediators should not be forced to disclose all their information in the event of breakdown. Trust in the mediator's discretion is an important driving force of the success of a mechanism.

More broadly, we show that the most important aspect of the optimal ADR-protocol is the management of the information structure in litigation post-breakdown. The optimal protocol imposes type-independent beliefs to minimize the potential gain a deviator can earn in the litigation game following a misreport. In addition, the protocol is asymmetric to reduce resource intensity in case of breakdown.

We demonstrate that the standard assumption of fixed, type-dependent outside options in mechanism design is not innocuous when the following two conditions are satisfied: (1) the mechanism cannot replace the underlying default game completely and (2) the actions chosen in the underlying game depend on player's beliefs. We show that the behavior of the players in the mechanism and those in the underlying game are interconnected. For the case of contests, we show that players invest less resources post-breakdown for extreme type distributions compared to a situation in which no resolution mechanism is present. For intermediate type-distributions, however, the post-mediation contest can also be more resource intensive.

Not claiming that the actual ADR-mechanisms we observe in reality are optimal, we want to note that our findings are in line with some observations on ADR. Its success rates are beyond 50% across cases and jurisdictions and mediation is considered to be informative when breaking down. In addition, one reason why mediation is perceived to be successful is its ability to not rely on publicly observable actions of the mediator, but allowing for private settlement negotiations.

Our findings provide several interesting directions for future research. First of all, the assumption that the mechanism designer has full-commitment could be relaxed to allow for third-party renegotiation. Especially when mediators compete for clients this seems reasonable. Further, extending the analysis to a setup of more than two players and possibly correlated types might add several interesting channels to the model. In addition, many conflicts evolve around a variety of battlefields on different subjects or points in time. If types are correlated over time this adds an additional signaling dimension which is interesting to analyze further. Finally, although minimizing court appearances is optimal given the public good properties of the legal system, it is less clear in other contest situations whether this is the most suitable objective. Although a richer model is needed to address such issues properly, we are confident that the results of this papers provide a first step towards analyzing these problems.

Appendix A

Appendix Chapter 2

A.1 Proof of Lemma 2.1

Take any $\tilde{\beta}$-feasible allocation ϕ that induces the play of the default game (i.e., $q_{\phi}^{A,P}(\hat{j},\hat{k}) > 0$ for some profile of reports). Agent type j's utility reads: $U_{\phi}^{A}(\hat{j},j,\tilde{\beta}) = \sum_{k=0}^{K}\tilde{\beta}_k(q_{\phi}^{A}(\hat{j},k)j + q_{\phi}^{A,P}(\hat{j},k)U_D^A(j,k) - t_{\phi}^{A}(\hat{j},k))$. Principal type k's utility reads: $U_{\phi}^{P}(\hat{k},k) = \int_0^J(q_{\phi}^{P}(j,\hat{k})\theta_k + q_{\phi}^{A,P}(j,\hat{k})U_D^P(j,k) - t_{\phi}^{P}(j,\hat{k}))f(j)dj$. We want to show that there exists ϕ' that does not involve the non-cooperative play of the auction, is $\tilde{\beta}$-feasible and leaves every bidder with the same payoff as ϕ.

For any $\hat{k} \in supp(\tilde{\beta})$ define ϕ' by $q_{\phi'}^{A}(\hat{j},\hat{k}) = q_{\phi}^{A}(\hat{j},\hat{k}) + q_{\phi}^{A,P}(\hat{j},\hat{k})\mathbb{1}(\hat{j} > \hat{k})$, $q_{\phi'}^{P}(\hat{j},\hat{k}) = q_{\phi}^{P}(\hat{j},\hat{k}) + q_{\phi}^{A,P}(\hat{j},\hat{k})\mathbb{1}(\hat{j} < \hat{k})$, $t_{\phi'}^{A}(j,k) = t_{\phi}^{A}(\hat{j},\hat{k}) + \mathbb{1}(\hat{j} > \hat{k})q_{\phi}^{A,P}(\hat{j},\hat{k})(\theta_{\hat{k}})$, $t_{\phi'}^{P}(\hat{j},\hat{k}) = t_{\phi}^{P}(\hat{j},\hat{k}) + \mathbb{1}(\hat{j} < \hat{k})q_{\phi}^{A,P}(\hat{j},\hat{k})(\hat{j})$ and $q_{\phi'}^{A,P}(\hat{j},\hat{k}) = 0$.

Observe, by construction $U_{\phi}^{A}(j,j,\tilde{\beta}) = U_{\phi'}^{A}(j,j,\tilde{\beta}) \ \ \forall j \in \Theta^A$, $U_{\phi}^{P}(k,k) = U_{\phi'}^{P}(k,k) \ \ \forall k \in supp(\tilde{\beta})$ and $q_{\phi'}^{A,P}(\hat{j},\hat{k}) = 0 \ \ \forall \hat{j},\hat{k} \in \Theta^A \times supp(\tilde{\beta})$.

Suppose that ϕ' is $\tilde{\beta}$-incentive-compatible for the principal. Then, $U_{\phi'}^{A}(\hat{j},j,\tilde{\beta}) = \sum_{k=0}^{K}\tilde{\beta}_k\{q_{\phi}^{A}(\hat{j},k)j - t_{\phi}^{A}(\hat{j},k) + \mathbb{1}(\hat{j} > k)q_{\phi}^{A,P}(\hat{j},k)(j-\theta_k)\} \leq U_{\phi}^{A}(\hat{j},j,\tilde{\beta}) \ \ \forall \hat{j} \in \Theta^A$, and with equality if $\hat{j} = j$. Since the right-hand side is maximized when $\hat{j} = j$ (by $\tilde{\beta}$-incentive compatibility of ϕ), ϕ' is $\tilde{\beta}$-incentive compatibility for the agent. A similar reasoning applies when reversing the roles of the principal with type in the support of $\tilde{\beta}$ and the agent.

Finally, consider a principal type k not in the support of $\tilde{\beta}$. Denote k's optimal report given ϕ by $\hat{k}(k)$. For any such type, augment ϕ' by a menu such that $(q_{\phi'}^{P}(\hat{j},k), q_{\phi'}^{A,P}(j,k), t_{\phi'}^{P}(\hat{j},k)) = (q_{\phi}^{P}(\hat{j},\hat{k}(k)) + q_{\phi}^{A,P}(\hat{j},\hat{k}(k))\mathbb{1}(k > \hat{j}), 0, t_{\phi}^{P}(\hat{j},\hat{k}(k)) + q_{\phi}^{A,P}(\hat{j},\hat{k}(k))(\hat{j})\mathbb{1}(k > \hat{j}))$. By construction, $U_{\phi'}^{P}(k,k) = U_{\phi}^{P}(\hat{k}(k),k)$. Moreover, the $\tilde{\beta}$-incentive constraints of those principal types in the support of $\tilde{\beta}$ are not altered, since ϕ is $\tilde{\beta}$-incentive-compatible: According to ϕ, each principal type k' in the support of $\tilde{\beta}$ can secure himself the utility $U_{\phi'}^{P}(\hat{k}(k),k')$ by reporting to be $\hat{k}(k)$ and submitting the bid θ_k in the SPA, whenever the mechanism allows him to do so. Since this is a (weakly) dominated strategy, it follows from $\tilde{\beta}$-incentive compatibility of ϕ that $U_{\phi'}^{P}(\hat{k}(k),k') \leq U_{\phi}^{P}(\hat{k}(k),k') \leq U_{\phi}^{P}(k',k') = U_{\phi'}^{P}(k',k')$.

We conclude that there is no loss of generality on focusing on outcome functions that do not involve the play of the default game, whenever evoking the concept of $\tilde{\beta}$-feasibility. Since the argument can easily be generalized to capture the case of safe allocations, we consider statement (ii) as proven.

A.2 Characterization of the RSW Allocation and the Proof of Lemma 2.2

Similar to proposition 2 in Maskin and Tirole, 1992, where the agent has no private information, the solution ϕ^{RSW} to the class of maximization problems $(P1)_{\check{k}}$ for $\check{k} \in \Theta^P$, can be found and thought of as solving the following sequence of linked maximization problems:

$$\max_{\phi(\cdot,0)} U^p_{\phi(\cdot,0)}(0) \qquad (\tilde{P}1)_0$$

such that $\forall j \in \Theta^a$:

$$(AIR)^j_0, (AIC)^j_0, (BB)_{j,0},$$

and for any $k = 1, ..., K$ and given $\phi(\cdot, 0), ..., \phi(\cdot, k-1)$,

$$\max_{\phi(\cdot,k)} U^p_{\phi(\cdot,k)}(k) \qquad (\tilde{P}1)_k$$

such that $\forall j \in \Theta^a$:

$$(AIR)^j_k, (AIC)^j_k, (BB)_{j,k},$$

$$(PIC)^+_{k-1} \qquad U^p_{\phi(\cdot,k-1)}(k-1) \geq U^p_{\phi(\cdot,k)}(k-1),$$

where each $(\tilde{P}1)_k$ defines an optimal - from principal type k's point of view - menu offer to the agent, satisfying $k-1$'s upward adjacent incentive constraint, $(PIC)^+_{k-1}$. Each $(\tilde{P}1)_k$ is subject to the agent's ex-post individual rationality and incentive constraint, given k.

We start with principal type $k = 0$. We construct the optimal menu $\phi^{RSW}(\cdot, 0)$ from his point of view. We continue with principal type $k = 1$. We construct the optimal menu $\phi^{RSW}(\cdot, 1)$ from his point of view, subject to the additional constraint that $k = 0$ does not prefer to offer $\phi^{RSW}(\cdot, 1)$ instead of $\phi^{RSW}(\cdot, 0)$, i.e., subject to $(PIC)^+_0$. Having determined $\phi^{RSW}(\cdot, 1)$, we proceed in the same fashion with the next higher principal type, and so on.

RSW Allocation. *There exists $\underline{K} \in \Theta^P \cup K+1$ and a sequence of increasing threshold types $\{j(k)\}_{k=0}^{\underline{K}-1}$ with $j(k) \in \Theta^A$ and $j(0) = 0$, such that ϕ^{RSW} is as follows:*

$\forall (j,k) \in \Theta$ *with* $k < \underline{K}$:

$$q^P_{RSW}(j,k) = \mathbb{1}[j \leq j(k)], \quad q^A_{RSW}(j,k) = 1 - q^P_{RSW}(j,k)$$

$$t^P_{RSW}(j,k) = \mathbb{1}[j \leq j(k)](-\theta_k) + (1 - \mathbb{1}[j \leq j(k)])(j(k) - \theta_k), \quad t^A_{RSW}(j,k) = -t^P_{RSW}(j,k)$$

$\forall (j,k) \in \Theta$ *with* $k \geq \underline{K}$:

$$q^P_{RSW}(j,k) = 1, \quad q^A_{RSW}(j,k) = 1 - q^P_{RSW}(j,k)$$

$$t^P_{RSW}(j,k) = \theta_{\underline{K}-1} - U^P_{RSW}(\underline{K}-1), \quad t^A_{RSW}(j,k) = -t^P_{RSW}(j,k)$$

To explain the form of the RSW allocation, let us describe the menus, induced by the sequence of maximization problems $(\tilde{P}1)_k$: Principal type 0 does not value the right. Since he is constraint by $(AIR)^j_0$, $\phi_{RSW}(\cdot, 0)$ is a degenerate menu that allots the right to every agent type at zero transfer payment.

A menu associated with principal type k, such that $0 < k < \underline{K}$, leaves the agent with a choice set consisting of two elements: The right to be the only bidder in the SPA and the commitment to stay out of the auction. Augmented to each element is a transfer. The construction of these transfers assures that every agent type above a threshold, $j(k)$, chooses the first element and every type below the threshold selects the second element. Imposing in addition that the agent's ex-post individual rationality-constraint has no slack, these transfers are uniquely determined as function of $j(k)$ and k. If the agent type is above the threshold, one might think of the transfer as price for the right. If the type is below the threshold, the transfer takes the role of a compensation, paid to the agent for the commitment to stay out of the auction. Via the binding resource-feasibility and budget-balanced constraint, the agent's choice in the menu associated with principal type k determines an expected allocation for the latter. The parameter $j(k)$ is chosen such that principal type $k-1$ indeed prefers the menu associated with his type and does not imitate k. That is, $j(k)$ is determined by principal's type $k-1$ upward adjacent incentive constraint at the interim level. The higher the threshold $j(k)$, the higher the probability with which the right is allotted to principal type k. Since the principal's upward adjacent incentive constraints are binding, the probability with which type k receives the right is inefficiently high.The resulting inefficient binary 0,1 allotment policy implies that the ex-post individual rationality-constraints of those agent types being weakly above $j(k)$ are satisfied with equality. Incentive compatibility in turn implies that the ex-post individual rationality-constraints for lower types than $j(k)$ is satisfied with strict inequality.

If we can construct menus in the above fashion for all principal types, we set $\underline{K} = K + 1$. If, however, for some sufficiently high principal type $\tilde{k}$ there does not exist a menu generated in the above manner that separates $\tilde{k}$ from his downward adjacent neighbor $\tilde{k}-1$, i.e., when choosing $j(\tilde{k}) = J$ and constructing the incentive-compatible transfers such that the agent's individual rationality-constraints have no slack -, we set $\underline{K} = \tilde{k}$. In this instance, we refer to the set of principal types weakly above $\underline{K}$ as the pool at the top. All types of the principal weakly above $\underline{K}$ offer the same degenerate menu. Upon acceptance, the agent commits to stay out of the auction in exchange for a transfer payment. This payment from the principal to the agent is chosen such that principal type $\underline{K} - 1$'s upward adjacent incentive constraint is satisfied with equality. As implication, every agent type's ex-post individual rationality-constraint with respect to the principal's types in the pool at the top holds with strict inequality.[1]

For completeness, note the optimality of the binary 0,1 allotment policy is implied by the regularity assumption on the agent's type distribution, i.e., assumption 1.

Outline of the proof We first state an algorithm that defines the RSW allocation completely. For later purpose, we continue by stating a different maximization problem $(P1)'$, being subject to the same set of constraints as any $(P1)_{\check{k}}$. We prove that the solution to $(P1)'$ solves every $(P1)_{\check{k}}$. We work with $(P1)'$ and show that the claimed solution ϕ^{RSW} solves $(P1)'$. In order to set up a tractable maximization problem, we make use of some ideas of Ledyard and Palfrey, 2007. By

[1] Referring to Maskin and Tirole, 1992, the existence of such a pool at the top corresponds to a failure of their sorting assumption.

relaxing the agent's incentive constraints and restating his individual rationality-constraints, we transform $(P1)'$ to a program with finitely many constraints. This new program is sufficiently well-behaved in that it admits a quasi-concave objective and a convex constraint-set. Moreover, the problem admits an interior point, ϕ^o, defined in the proof of lemma 2.3 in A.3. Thus, we are able to apply the standard Lagrangian-methodology. We proceed by constructing multipliers being consistent with complementary slackness, given the assumed optimality of the RSW allocation. Given these multipliers, we show that $q_{RSW}(\cdot,\cdot)$ maximizes the Lagrangian function (point-wise). By the imposed regularity condition on the agent's distribution, i.e., assumption 1, the ignored monotonicity condition is satisfied at the optimum.[2]

Algorithm.

- *Step 0: Start with $k = 1$. Note that $(q^P_{RSW}(0), t^P_{RSW}(0))$ and thus $U^P_{RSW}(0)$ are determined by the RSW allocation stated above. Go to step 2.*

- *Step 1:*

  For any $k > 1$, note that $(q^P_{RSW}(k-1), t^P_{RSW}(k-1))$ are determined up to $j(k-1)$, according to the form of the RSW allocation. Hence, so is $U^P_{RSW}(k-1, k-1) = q^P_{RSW}(k-1)\theta_{k-1} - t^P_{RSW}(k-1)$. $j(k-1)$ is determined by the algorithm starting with $k-1$.

- *Step 2:*

  If $k > J$, set $j(k) = J$

  If $k < J$,compute the set Ω_k, where

  $\Omega_k := \{j \in \Theta^A | U^P_{RSW}(k-1, k-1) \geq F(j)\theta_{k-1} + (1-F(j))\theta_k - F(j)\max\{j - \theta_k, 0\}\}$

  If Ω_k is empty, set $\underline{K} = k$ and stop. In this instance set $j(k) = J \ \ \forall k \geq \underline{K}$.

  Otherwise define $j(k) := \min \Omega_k$.

- *Step 3:*

  Given $j(k)$, $U^P_{RSW}(k,k)$ is completely determined by the formula of the RSW allocation.

- *Step 4:*

  If $k < K$, go back to step 1 and start with $k+1$.

  If $k = K$ go to step 5.

- *Step 5:*

  Set $\underline{K} = K + 1$.

[2] The way we solve the K+1 maximization problems that define the RSW allocation seems rather elaborate. It is taken, since we make use of the characterization of the solution to $(P1)'$ in terms of the multipliers when proving that the RSW allocation is undominated for some $\tilde{\beta}$ and welfare weights.

Some remarks are in order: First, $\underline{K}$ takes a value weakly smaller than K iff the algorithm stops in stage 2 for some $k \in \Theta^P$. A sufficient condition for this event is $\theta_K \geq J$. In this instance, Ω_K is empty and thus $\underline{K} \leq K$. Second, note that $j(k)$ is a weakly increasing sequence. Third, for all $k < \underline{K}$ it holds that $j(k) > \theta_k$. This follows by construction: Suppose $j(k) \leq \theta_k$, then $\max\{j(k) - \theta_k, 0\} = 0$. Therefore, the right-hand side of the constraint in Ω_k takes a value strictly above θ_{k-1}. Since $U^P_{RSW}(\hat{k} = k-1, k-1) \leq \theta_{k-1}$, this is a contradiction to the fact that $j(k) \in \Omega_k$.

Introducing auxiliary program $(P1)'$ In order to prove that ϕ^{RSW} solves any problem $(P1)_{\check{k}}$, we next show that this solution satisfies the optimality conditions of program $(P1)'$.

$$\max_{\phi} \sum_{k=0}^{K} U^P_\phi(k) \qquad (P1)'$$

such that $\forall j, k \in \Theta$:

$$(AIR)^j_k, (AIC)^j_k, (PIC)_k, (BB)_{j,k}$$

We start by relaxing the constraints:

Step 0: (PIC)

We replace (PIC) by $(PIC)^+$. That is, we impose the constraint:

$$(PIC)^+_k \qquad U^P_\phi(k,k) \geq U^P_\phi(k+1,k) \;\; \forall k \in \{0,1,...,K-1\}.$$

Step 1: $(AIC)_k$

Recall that given any k, agent's utility takes the form:

$$U^A_\phi(\hat{j}, j, k) = q^A_\phi(\hat{j}, k) j - t^A_\phi(j,k).$$

Note, $\frac{\partial U^A_\phi(\hat{j},j,k)}{\partial j} \leq 1$, for all $q^A_\phi(\hat{j},k), t^A_\phi(\hat{j},k)$. By Theorem 2 of Milgrom and Segal (2002) we thus know that the incentive-compatible form of the utility is given by:

$$U^A_\phi(j,k) = \int_0^j q^A_\phi(v,k)dv + U^A_\phi(0,k),$$

where $U^A_\phi(0,k) := q^A_\phi(0,k)0 + t^A_\phi(0,k)$ is the utility of the lowest agent type. Moreover, $U^A_\phi(j,k)$ is absolutely continuous in j.

If in addition $q^A_\phi(j,k)$ is non-decreasing in j, $(AIC)_k$ is satisfied.

Step 2: $(AIR)_k$

By the above necessary condition, individual rationality reads (given any k):

$$\int_0^j q^A_\phi(v,k)dv + U^A_\phi(0,k) - U^A_D(j,k) \geq 0.$$

Define

$$\tilde{J}(k) := arg \min_{j \in \Theta^A} \{ \int_0^j (q^A_\phi(v,k))dv - U^A_D(j,k)\},$$

which might be a set of (possibly infinitely many) types.

We guess that $J \in \tilde{J}(k)$, which will later be verified. By this guess and the definition of $\tilde{J}(k)$ it immediately follows that $U^A_\phi(J,k) \geq U^A_D(J,k)$ implies $(AIR)^j_k$ is satisfied for all j.
We replace $(AIR)_k$ by $U^A_\phi(J,k) - U^A_D(J,k) \geq 0$ and denote the relevant multiplier by ρ_k.
Step 3: $(BB)_{j,k}$
We first relax $(BB)_{j,k}$ to

$$(BB)_k \qquad \mathbb{E}_j t^A_\phi(j,k) + t^P_\phi(k) \geq 0 \ \ \forall k \in \Theta^P.$$

Second, we make use of step 1 to obtain the following representation for the transfer of any j given any k:

$$\begin{aligned} t^A_\phi(j,k) &= q^A_\phi(j,k)j - U^A_\phi(j,k) \\ &= q^A_\phi(j,k)j - \int_0^j q^A_\phi(v,k)dv - t^A_\phi(0,k). \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}_j t^A_\phi(j,k) &= \int_0^J q^A_\phi(j,k)jf(j)dj - \int_0^J (\int_0^j q^A_\phi(v,k)dv)f(j)dj - t^A_\phi(0,k) \\ &= \int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj - t^A_\phi(0,k), \end{aligned}$$

where the last step follows from partial integration.
We thus replace $(BB)_{j,k}$ by:

$$\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj - t^A_\phi(0,k) + t^P_\phi(k).$$

We denote the multiplier on this constraint by δ_k.
Step 4: Lagrangian
We are ready to state the Lagrangian objective:

$$\max \mathcal{L} = \sum_{k=0}^{K} U^P_\phi(k) + \sum_{k=0}^{K} \rho_k[\int_0^J (q^A_\phi(v,k))dv + t^A_\phi(0,k) - U^A_D(J,k)]$$

$$+ \sum_{k=0}^{K-1} v_{k,k+1}[U^P_\phi(k,k) - U^P_\phi(k+1,k)] + \sum_{k=0}^{K} \delta_k[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj - t^A_\phi(0,k) + t^P_\phi(k)],$$

where $v_{k,k+1}$ and $v_{k-1,k}$ denote the multiplier on the $(PIC)^+$ constraints.
Complementary slackness requires:

$$\rho_k[\int_0^J (q^A_\phi(v,k))dv + t^A_\phi(0,k) - U^A_D(J,k)] = 0 \ \ \forall k \in \Theta^P,$$

$$v_{k,k+1}[U^P_\phi(k,k) - U^P_\phi(k+1,k)] = 0 \ \ \forall k \in \Theta^P \setminus K,$$

$$\delta_k[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj - t^A_\phi(0,k) + t^P_\phi(k)] = 0 \ \ \forall k \in \Theta^P.$$

The choice variables are the following:

$$\{t^A_\phi(0,k), t^P_\phi(k), q^A_\phi(\cdot,k), q^P_\phi(\cdot,k)\}_{k=0}^K$$

Step 5: Necessary conditions for an optimum

As first step, we use the first order conditions of the choice variables $t^A_\phi(0,k)$ and $t^P_\phi(k)$ to construct non-negative multipliers being consistent with complementary slackness under the hypothesis that $\phi = \phi^{RSW}$ solves the problem.
The derivative with respect to $t^A_\phi(0,k)$ satisfies:

$$\rho_k - \delta_k = 0 \tag{A.1}$$

By the claimed optimality of ϕ^{RSW} it follows that $\rho_k = 0$ (and thus $\delta_k = 0$) whenever $k \geq \underline{K}$. If $k < \underline{K}$ the claimed optimality implies that $\rho_k = \delta_k \geq 0$.
The derivative with respect to $t^P_\phi(k)$ satisfies:

$$-1 - v_{k,k+1} + v_{k-1,k} = -\delta_k \tag{A.2}$$

Summing (A.2) over k yields

$$\sum_{k=0}^{K} 1 = \sum_{k=0}^{\underline{K}-1} \delta_k := \delta.$$

Iterating backwards on (A.1) motivates the following choice of the multipliers:

$$v_{k-1,k} := \sum_{v=k}^{K} 1 \ \text{ if } \ k \geq \underline{K},$$

$$v_{k-1,k} := (\sum_{v=0}^{K} 1 - \sum_{v=0}^{k-1} 1) - (\delta - \sum_{v=0}^{k-1} \delta_v) = \sum_{v=0}^{k-1} \delta_v - \sum_{v=0}^{k-1} 1 \ \text{ if } \ k < \underline{K}.$$

Note, for $k = 1$ it follows that $v_{0,1} = \delta_0 - 1$.
Non-negativity implies: $v_{k-1,k} = \sum_{k=0}^{k-1} \delta_k - \sum_{k=0}^{k-1} 1 \geq 0 \ \ \forall k > 0$.
We thus face the following restrictions on the sequence $\{\delta_k\}_{k=0}^K$ with $\delta_k = 0 \ \forall k \geq \underline{K}$:

$$(i) \quad \delta_k \geq 0, \quad (ii) \quad v_{k-1,k} = \sum_{v=0}^{k-1} (\delta_v - 1) \geq 0, (iii) \quad \sum_{k=0}^{K} \delta_k = \sum_{k=0}^{K} 1 \tag{F1}$$

Step 6: Manipulating the Lagrangian terms: $(PIC)^+$

We aim to manipulate the Lagrangian objective in order to derive an expression that allows us to solve the maximization problem point-wise. As first step, consider the following expression:

$$A := \sum_{k=0}^{K} U^P_\phi(k) + \sum_{k=0}^{K-1} v_{k,k+1}[U^P_\phi(k,k) - U^P_\phi(k+1,k)] + \sum_{k=0}^{K} \delta_k[t^P_\phi(k)].$$

Note that

$$v_{k,k+1}[U^P_\phi(k,k) - U^P_\phi(k+1,k)] = -v_{k,k+1}[\theta_k(q^P_\phi(k+1) - q^P_\phi(k)) - (t^P_\phi(k+1) - t^P_\phi(k))].$$

Substituting in A and using the first order conditions with respect to $t^P_\phi(k)$, (A.2), leads:

$$A=\sum_{k=0}^{K}\theta_k q^P_\phi(k)-\sum_{k=0}^{K-1}v_{k,k+1}[\theta_k(q^P_\phi(k+1)-q^P_\phi(k))-(t^P_\phi(k+1)-t^P_\phi(k))]+\sum_{k=0}^{K}(\delta_k-1)[t^P_\phi(k)]$$

$$=\sum_{k=0}^{K}\theta_k q^P_\phi(k)-\sum_{k=0}^{K-1}v_{k,k+1}[\theta_k(q^P_\phi(k+1)-q^P_\phi(k))].$$

Applying some algebra manipulations we can reformulate A as:

$$A=\sum_{k=1}^{K-1}q^P_\phi(k)[\theta_k(1+v_{k,k+1})-v_{k-1,k}\theta_{k-1}]+q^P_\phi(K)\theta_K+(1+v_{0,1})q^P_\phi(0)\theta_0.$$

Again, by the first order conditions, (A.2), and the assumed form of the multipliers:

$$q^P_\phi(k)[\theta_k(1+v_{k,k+1})-v_{k-1,k}\theta_{k-1}]=q^P_\phi(k)[\delta_k\theta_k+(\theta_k-\theta_{k-1})v_{k-1,k}]$$

$$=q^P_\phi(k)[\delta_k\theta_k+(\theta_k-\theta_{k-1})\sum_{k=0}^{k-1}(\delta_k-1)]\ \text{ if }\ 0<k<\underline{K},$$

and

$$q^P_\phi(k)[\theta_k(1+v_{k,k+1})-v_{k-1,k}\theta_{k-1}]=q^P_\phi(k)[(\theta_k-\theta_{k-1})v_{k-1,k}]$$

$$=q^P_\phi(k)(\theta_k-\theta_{k-1})\sum_{k=k}^{K}1\ \text{ if }\ K>k\geq\underline{K}.$$

Merging these increments leads:

$$A=\sum_{k=1}^{\underline{K}-1}q^P_\phi(k)[\delta_k\theta_k+(\theta_k-\theta_{k-1})\sum_{k=0}^{k-1}(\delta_k-1)]+\sum_{k=\underline{K}}^{K-1}[q^P_\phi(k)(\theta_k-\theta_{k-1})\sum_{k=k}^{K}1]+q^P_\phi(K)\theta_K+\delta_0 q^P_\phi(0)\theta_0.$$

Step 7: Manipulation of the Lagrangian objective, continued

We want to use the previous results to reformulate the Lagrangian. First, observe that $\mathcal{L}=A+B$, with B being defined as:

$$B:=\sum_{k=0}^{K}\rho_k[\int_0^J(q^A_\phi(v,k))dv+t^A_\phi(0,k)-U^A_D(J,k)]$$

$$+\sum_{k=0}^{K}\delta_k[\int_0^J\{(j-\frac{1-F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj-U^A_\phi(0,k)].$$

We can reformulate B:

$$B=\sum_{k=0}^{\underline{K}-1}\delta_k[\int_0^J\{(j+\frac{F(j)}{f(j)})q^A_\phi(j,k)\}f(j)dj-U^A_D(J,k)].$$

$\mathcal{L}$ now reads:

$$\mathcal{L} = \sum_{k=1}^{\underline{K}-1} \int_0^J \{q_\phi^P(j,k)[\theta_k + (\theta_k - \theta_{k-1})\frac{\sum_{k=0}^{k-1}(\delta_k - 1)}{\delta_k}] + (j + \frac{F(j)}{f(j)})q_\phi^A(j,k)\}\delta_k f(j)dj$$

$$-\sum_{k=1}^{\underline{K}-1} \delta_k U_D^A(J,k) + \sum_{k=\underline{K}}^{K-1}[q_\phi^P(k)(\theta_k - \theta_{k-1})\sum_{k=k}^{K} 1] + q_\phi^P(K)\theta_K$$

$$+q_\phi^P(0)\theta_0 + \int_0^J (j + \frac{F(j)}{f(j)})q_\phi^A(j,0)f(j)\}dj$$

$$+F(0)\sum_{k=1}^{\underline{K}-1} q_\phi^P(k,0)[\delta_k\theta_k + (\theta_k - \theta_{k-1})\sum_{k=0}^{k-1}(\delta_k - 1)]$$

Step 8: Optimal allotment policy

For given $\{\delta_k\}_{k=0}^{\underline{K}-1}$ the claimed optimality of the RSW allocation follows if q_{RSW} maximizes $\mathcal{L}$ point-wise. We thus need to find $\{\delta_k\}_{k=0}^{\underline{K}-1}$, satisfying the above stated optimality conditions, such that the claimed allotment policy satisfies the feature of a point-wise maximizer, say q_{ϕ^*}.

For any $k \geq \underline{K}$ it is straightforward to observe that $q_{\phi^*}^P(k) = 1$. Moreover, since $\theta_0 = 0$ it directly can be seen from the shape of $\mathcal{L}$ that $q_{\phi^*}^P(j,0) = 0$ and $q_{\phi^*}^A(j,0) = 1$ and $q_{\phi^*}^P(0,k) = 1$ and $q_{\phi^*}^A(0,k) = 0$.

Now consider $k \in \{1, ..., \underline{K} - 1\}$. We want to show that we can find $\{\delta_k\}_{k=0}^{\underline{K}-1}$ such that ϕ^{RSW} maximizes $\mathcal{L}$ point-wise. We are free to choose any $\{\delta_k\}_{k=0}^{\underline{K}-1}$ being consistent with (F1).

For any k, such that $\underline{K}-1 \geq k > 0$ we must find a sequence of multipliers, $\sum_{k=0}^{K} 1 = \sum_{k=0}^{K} \delta_k$ satisfying the optimality requirement:

$$[\theta_k + (\theta_k - \theta_{k-1})\frac{\sum_{v=0}^{k-1}(\delta_v^* - 1)}{\delta_k}] = j(k) + \frac{F(j(k))}{f(j(k))},$$

where we treat $j(k)$ as exogenously given.[3]

If this sequence is consistent with (F1), it follows that $\sum_{k=0}^{K} 1 = \sum_{k=0}^{K} \delta_k$. This allows us to rewrite:

$$\sum_{v=0}^{k-1}(\delta_v - 1) = \sum_{v=k}^{K}(1 - \delta_v).$$

We apply the following approach: We guess the existence of a solution to the problem, say $\{\delta_k^*\}_{k=0}^{\underline{K}-1}$, that satisfies (F1). Afterwards, we apply backward induction to first verify that any such δ_k^*, with $0 < k \leq \underline{K} - 1$, satisfies the optimality requirement and features $0 < \delta_k^* < [\sum_{v=k+1}^{K}(1 - \delta_v^*)] + 1$. Finally, we construct δ_0^*, and verify that all conditions in (F1) are satisfied.

We now apply the induction procedure.

We argue that, for any k and given $\{\delta_v^*\}_{v=k+1}^{\underline{K}-1}$, we can find δ_k such that:

$$[\theta_k + (\theta_k - \theta_{k-1})\frac{\sum_{v=k+1}^{K}(1 - \delta_v^*) + (1 - \delta_k)}{\delta_k}] = j(k) + \frac{F(j(k))}{f(j(k))}.$$

[3] Recall $j(k) := arg\min\{j \in \Theta^A | q_{RSW}^A(j,k) = 1\}$

Since for fixed $\{\delta_v^*\}_{v=k+1}^{\underline{K}-1}$ the left-hand side is continuous in δ_k, the intermediate value theorem applies. First, when δ_k is arbitrarily small, the left-hand side goes to (plus) infinity. This follows from the guess that the requirements are satisfied, i.e., $\sum_{v=k+1}^{K}(1-\delta_v^*) > 0$. When $\delta_k = \bar{\delta_k} := [\sum_{v=k+1}^{K}(1-\delta_v^*)] + 1$ the left-hand side is strictly smaller than the right-hand side (this follows since $\theta_k < j(k)$). Moreover, the left-hand side strictly decreases in δ_k. We thus conclude that for any k such that $\underline{K} - 1 \geq k > 0$ the solution to the above equation, δ_k^*, features: $0 < \delta_k^* < [\sum_{v=k+1}^{K}(1-\delta_v^*)] + 1$.
As last step, consider $k = 0$. We determine δ_0^* such that $\sum_{k=0}^{K}\delta_k = \sum_{k=0}^{K} 1$. Let

$$\delta_0^* := 1 + \sum_{k=1}^{\underline{K}-1}(1-\delta_k^*) + \sum_{k=\underline{K}-1}^{K} 1.$$

Moreover, $\delta_0^* > 1$ because $\delta_1^* - 1 - \sum_{k=2}^{\underline{K}-1}(1-\delta_k^*) < 0$, by the features of δ_k^*. We are now ready to verify the conditions stated in (F1). First, observe that

$$\delta = \sum_{k=0}^{\underline{K}-1}\delta_k^* = \sum_{k=0}^{K} 1,$$

by construction of δ_0^*. This proves (F1) (iii).
Second,

$$\sum_{v=0}^{k-1}(\delta_k^* - 1) \geq 0 \;\; \forall k, \quad \iff \sum_{k}^{K}(1-\delta_k^*) \geq 0 \;\; \forall k,$$

where the first step follows from the choice of δ_0^* and the last inequality from the feature of $\delta_k^* < \sum_{v=k+1}^{K}(1-\delta_v^*) + 1$. This verifies (F1) (ii).
Third, note that

$$\delta_k^* \geq 0$$

is obviously satisfied for all $k > 0$ and therefore F1 (i) follows.
Step 9: Optimality of point-wise maximization
We constructed the multiplier such that the RSW allocation is consistent with point-wise maximization of $\mathcal{L}$. It is left to argue that point-wise maximization leads the optimum. This is true because $j + \frac{F(j)}{f(j)}$ is non-decreasing in j and $j(k)$ is weakly increasing in k.
The next Lemma completes the proof of Lemma 2.2.

Lemma A.1. *Suppose ϕ' is a solution to $(P1)'$. Then ϕ' is a solution to every $(P1)_{\check{k}}$.*

Proof. Suppose to the contrary that ϕ' does not solve every $(P1)_{\check{k}}$. That is, there exists at least one principal type k' such that $U^P_{RSW}(k') > U^P_{\phi'}(k')$. Recall that both maximization problems are subject to the same set of constraints. Construct a new allocation, ϕ'', by setting $\phi''(\cdot,k) = \phi_k'''(\cdot,k)$ for all k, where $\phi_k''' := \arg\max_{\phi\in\{\phi^{RSW},\phi'\}} U^P_\phi(k,k)$. I.e., let each type of the principal choose whether he wants his allocation to be determined by the agent's choice in the menu $\phi_{RSW}(\cdot,k)$ or in $\phi'(\cdot,k)$. Note that ϕ'' satisfies the constraints of problem$(P1)'$: the agent's constraints are not altered, since they are satisfied at the ex-post level according to both ϕ^{RSW} and ϕ'. Suppose the principal's incentive constraint is violated for some type k''. Then, there must exists some $\hat{k} \neq k'' \in \Theta^P$ such that one of the following cases is true:

$$\max\{U^P_{RSW}(k'', k''), U^P_{\phi'}(k'', k'')\} < U^P_{RSW}(k''', k'')$$

$$\max\{U^P_{RSW}(k'', k''), U^P_{\phi'}(k'', k'')\} < U^P_{\phi'}(k''', k'')$$

This is a contradiction to the fact that both ϕ^{RSW} and ϕ' satisfy the principals incentive constraints.

We conclude that ϕ'' is feasible. By construction, we know that ϕ'' implies a weakly higher value of problem $(P1)'$ than ϕ'. By the assumption on the existence of k', we know that ϕ'' implies a strictly higher value of problem $(P1)'$ than ϕ'. Hence, ϕ' cannot solve problem $(P1)'$.

□

A.3 Proofs of Lemma 2.3 and Lemma 2.4

We verify that corollary 3 of Mylovanov and Tröger, 2015 applies to our setting. For this, we need to show that every principal type can implement his best safe payoff. That is, we need to verify that there exists an interior point (Step 1), that is, an allocation satisfying all defining constraints of a safe allocation with strict inequality. Given this, it follows that when the principal proposes a direct revelation mechanism equal to any convex combination between the allocation implementing his best safe payoff and the interior point, the agent accepts the mechanism and it features a unique equilibrium play. Afterwards, in Step 2 we verify that the secured payoff is implemented by the RSW allocation. In the course of this step, we also verify that the assumptions in Mylovanov and Tröger, 2015 are satisfied implying that there methodology applies.

Step 1

We first verify that there exists an allocation, say ϕ^o, that is safe with strict inequality. Denote by $\mu, \alpha > 0$ two parameters that are specified later. Then, ϕ^o takes the form:

$\forall j \in \Theta^A$:

$$(q^A_{\phi^o}(j,k), t^A_{\phi^o}(j,k)) = \Big\{ \; (\alpha F(j), \alpha(jF(j) - \textstyle\int_0^j F(v)dv) - J) \quad \forall k \in \Theta^P.$$

If $k = 0$:

$$(q^P_{\phi^o}(j,k), t^P_{\phi^o}(j,k)) = \Big\{ \; (0, 2J) \quad \forall j \in \Theta^A \,.$$

$\forall k > 0:$

$$(q^P_{\phi^o}(j,k), t^P_{\phi^o}(j,k)) = \Big\{ \; (\mu B(k), -U^P_{\phi^o}(k-1,k) + \tfrac{\theta_k + \theta_{k-1}}{2}\mu B(k)) \quad \forall j \in \Theta^A \,.$$

We choose $\mu, \alpha > 0$ sufficiently small, such that $\mu + \alpha < 1$.

Observe that ϕ^o is resource feasible and budget balanced since $q^A_{\phi^o}(j,k) + q^P_{\phi^o}(j,k) \leq \alpha + \mu < 1$ and $t^A_{\phi^o}(j,k) + t^P_{\phi^o}(j,k) \geq -J + 2J > 0 \quad \forall (j,k) \in \Theta$.

Moreover, ϕ^o creates an interior point: Firstly, $U^A_{\phi^o}(j,k) - U^A_D(j,k) = \alpha(\int_0^j F(v)dv) + J - \max\{0, j - \theta_k\} > 0$ for any $(j,k) \in \Theta$. Hence, $(AIR)^j_k$ is satisfied with strict inequality. Secondly, $(PIC)_k$ is satisfied with strict inequality, since by construction both the principal's upward and downward incentive constraints are satisfied with inequality. Finally, $(AIC)_k^j$ is satisfied with strict inequality, since $U^A_{\phi^o}(\hat{j}, j, k)$ is convex in $\hat{j}$ and has a unique maximum at $\hat{j} = j$.

Step 2

In order to apply Mylovanov and Tröger (2015), we introduce some further notation.

Define $H(k, \tilde{\beta})$ as the set of payoffs of type k, that is:

$$H(k, \tilde{\beta}) := \{U^P_\phi(k) | \exists \phi : (AIC)^j_{\tilde{\beta}}, (AIR)^j_{\tilde{\beta}}, (PIC)_k \;\; \forall j \in \Theta^A, \forall k \in supp(\tilde{\beta})\}.$$

Because of the quasilinear environment, $H(k, \tilde{\beta})$ is a convex set. Denote by U^p_ϕ the principal's payoff vector, with K+1 elements. Element $k+1$ corresponds to principal type k's payoff $U^P_\phi(k)$. Define the set $H(\tilde{\beta})$,

$$H(\tilde{\beta}) := \{U^P_\phi | \exists \phi : (AIC)^j_{\tilde{\beta}}, (AIR)^j_{\tilde{\beta}}, (PIC)_k \;\; \forall j \in \Theta^A, \forall k \in supp(\tilde{\beta})\}.$$

By the revelation principle, the continuation game starting after a proposal of any $m \in M$ together with fixed $\tilde{\beta}$ induces a $\tilde{\beta}$-feasible allocation. Let us denote the induced principal payoff vector by U^P. The k+1'th element, $U^P(k)$, denotes the payoff of type k. Hence, $U^P \in H(\tilde{\beta})$.

Moreover, let $M(\tilde{\beta}, m)$ denote the set of those payoff vectors being induced by the play of the continuation game subsequent to the principal's proposal of game form m with belief $\tilde{\beta}$ about the principal. In addition, define the set $\mathcal{K}$ such that $(\tilde{\beta}, U^P) \in \mathcal{K}$ if and only if $\tilde{\beta} \in B_0$ and $U^P \in M(\tilde{\beta}, m)$ for some $m \in M$.

We first verify that $\mathcal{K}$ is compact. Take a convergent sequence of beliefs $\tilde{\beta}_n \to \tilde{\beta}^*$ and a convergent sequence of payoff vectors such that $U^P_n \to U^P_*$ and $U^P_n \in H(\tilde{\beta}_n)$ for all n. For any element n, there exists a corresponding allocation ϕ_n that is $\tilde{\beta}_n$-feasible. We choose a convergent subsequence such that $\phi_n(j,k) \to \phi_*(j,k)$ for all $(j,k) \in \Theta$. As $H(\tilde{\beta}_n)$ is defined with respect to inequalities, it must be that $U^P_* \in H(\tilde{\beta}^*)$.

Next, we introduced the secured payoff defined by Mylovanov and Tröger (2015). For all $k \in \Theta^P$ the secured payoff, $\underline{U}^P(k)$, is defined as:

$$\underline{U}^P(k) := \underset{m \in M}{sup} \min_{\tilde{\beta} \in B, U^P \in M(\tilde{\beta}, m)} U^P_m(k).$$

Denote by $\underline{U}^P$ the secured-payoff vector, with the secured payoffs as elements.

We want to show that $U^P_{RSW}(k) = \underline{U}^P(k)$.

First, observe that $U^P_{RSW}(k) \leq \underline{U}^P(k)$: Suppose that $U^P_{RSW}(k) > \underline{U}^P(k)$. Consider the following direct revelation mechanism implementing ϕ^{RSW} with probability approaching 1 and ϕ^o with strictly positive probability approaching zero. By the features of ϕ^o, it follows that this direct revelation mechanism implements an allocation, ϕ', that is feasible for all beliefs. ϕ' induces principal type k payoff arbitrary close to U^P_{RSW}. This contradicts the definition of $\underline{U}^P(k)$.

Moreover, note that $U^P_{RSW}(k) \geq \underline{U}^P(k) \;\; \forall k \in \Theta^P$: Suppose $\underline{U}^P(k) > U^P_{RSW}(k)$ for some k. As initial step, observe $\underline{U}^P \in H(k, \tilde{\beta}) \;\; \forall \tilde{\beta}$ from the definition of the secured payoff: $\underline{U}^P(k) > U^P_\phi(k)$ for some $\tilde{\beta}'$, and all $\tilde{\beta}'$-feasible allocations implies a contradiction to the fact that $\underline{U}^P(k)$ is defined as the value of a minimization problem with respect to the belief. $\underline{U}^P(k) < U^P_\phi(k)$ for some $\tilde{\beta}'$, and all $\tilde{\beta}'$-feasible allocations cannot be the case since by hypothesis $U^P_{RSW}(k) < \underline{U}^P(k)$ and ϕ^{RSW} is $\tilde{\beta}$-feasible for all $\tilde{\beta}$. Since $H(k, \tilde{\beta})$ is a convex set, the claim follows. We thus note that there exists a payoff that is feasible for all beliefs. The revelation principle implies:

$$\underline{U}^P(k) \in \cap_{\tilde{\beta}\in B_0} H(k,\tilde{\beta}) = \{U^P_\phi(k) | \exists \phi : (AIC)^j_k, (AIR)^j_k, (PIC)_k \ \ \forall (j,k) \in \Theta\},$$

where the last equality follows from the fact that for any j and $\tilde{\beta}$, $U^A_D(j,\tilde{\beta}) = \sum_{k=0}^K \tilde{\beta}_k U^A_D(j,k)$. The hypothesis $\underline{U}(k) > U^P_{RSW}(k)$ thus implies a contradiction: There exists a game form that is feasible in $(P1)_k$ and leads a higher payoff than the solution value of this program.

A.4 Proof of Lemma 2.5

We show that there exist welfare weights ($\{z^*_k\}^K_{k=0}$ with $z^*_k > 0$) and a belief about the principal's types, β^*, such that the RSW allocation solves $(P2)$. We subsume those principal types not being in the support of β^* in the set O. I.e., $O := \{k \in \Theta^P | \beta^*_k = 0\}$. Whenever the pool at the top exists, we set $O = \{\underline{K}, ..., K\}$ and else $O = \emptyset$. We relax problem $(P2)$ by replacing the principal's global incentive constraints with local, upward adjacent constraints. The resulting program ,$(P2)'$, takes the following form:

$$\max_\phi \sum_{k=0}^K z^*_k U^P(k) \qquad (P2)'$$

such that $\forall j \in \Theta^A$ and $\forall k \in \{0, ..., \underline{K}-2\}$:

$$(AIR)^j_{\beta^*}, (AIC)^j_{\beta^*},$$

$$(PIC)^+_k \qquad U^P_\phi(k,k) \geq U^P_\phi(k+1,k),$$

for $k = \underline{K} - 1$:

$$(PIC)^O_{\underline{K}-1} \qquad U^P_\phi(\underline{K}-1, \underline{K}-1) \geq U^P_\phi(\hat{k}, \underline{K}-1) \quad \forall \hat{k} \in O,$$

and

$$\forall (j,k) \in \Theta \quad (BB)_{j,k}.$$

$(PIC)^O_{\underline{K}-1}$ imposes that principal type $\underline{K}-1$ does not prefer the menu being associated with a type of the principal in O to $\phi(\cdot, \underline{K}-1)$. Because O consists of those types being strictly above $\underline{K}-1$, this constraint is the local version of the constraint that no principal type $k' \in \Theta^P \setminus O$ does prefer $\phi(\cdot, k \in O)$ to $\phi(\cdot, k')$. Whenever the solution to the relaxed program induces a monotonic allotment policy, the imposed local constraints are sufficient for β^*-incentive compatibility at the optimum and the solution to $(P2)'$ solves $(P2)$.

We first relax the constraints of $(P2)'$. Afterwards, we construct the Lagrangian multipliers. Finally, we argue that there exists $\tilde{\beta}$ and $\{z^*_k\}^K_{k=0}$ such that ϕ^{RSW} solves the relaxed version of $(P2)''$. If O is empty, then all of the below still holds true. In this instance variables and arguments involving a type $k \in O$ are void.

Step 1: $(AIC)_{\beta^*}$

We work with the necessary condition and later verify that the sufficient condition for incentive compatibility is satisfied.

Note, $\frac{\partial U^A_\phi(\hat{j},j,\beta^*)}{\partial j} \leq 1$, for all $q^A_\phi(\hat{j},\beta^*), t^A_\phi(\hat{j},\beta^*)$.

By Theorem 2 of Milgrom and Segal, 2002 we thus know that the incentive-compatible form of the utility is given by:

$$U^A_\phi(j,\beta^*) = \int_0^j q^A_\phi(v,\beta^*)dv + U^A_\phi(0,\beta^*),$$

where $U^A_\phi(0,\beta^*) = q^A_\phi(0,\beta^*)0 + t^A_\phi(0,\beta^*)$ is the utility of the lowest agent type.

Moreover, $U^A_\phi(j,\beta^*)$ is absolutely continuous in j.

Step 2: $(AIR)_{\tilde{\beta}}$

By the above necessary condition, individual rationality reads (given any k):

$$\int_0^j (q^A_\phi(v,\beta^*)dv + t^A_\phi(0,\beta^*) - U^A_D(j,\beta^*) \geq 0.$$

Define

$$\tilde{J}^{\beta^*} := arg \min_{j\in\Theta^A}\{\int_0^j (q^A_\phi(v,\beta^*)dv - U^A_D(j,\beta^*)\},$$

which might be a set of (possibly infinitely many) types.

We guess that $\tilde{J}^{\tilde{p}} = J$, and verify it later. By this guess and the definition of $\tilde{J}^{\beta^*}$ it immediately follows that $U^A_\phi(J,\beta^*) \geq U^A_D(J,\beta^*)$ implies $(AIR)_{\beta^*}$ for all j.

We replace $(AIR)_{\beta^*}$ by $U^A_\phi(J,\beta^*) - U^A_D(J,\beta^*) \geq 0$ and denote the relevant multiplier by ρ.

Step 3: $(BB)_{j,k}$

We first relax $(BB)_{j,k}$ to

$$(BB)_k \qquad \mathbb{E}_j t^A_\phi(j,k) + \mathbb{E}_k t^P_\phi(k) \geq 0 \;\; \forall k \in \Theta^P.$$

Making use of step 1 we obtain the following representation for the transfer of any j:

$$t^A_\phi(j,\beta^*) = q^A_\phi(j,\beta^*)j - U^A_\phi(j,\beta^*)$$
$$= q^A_\phi(j,\beta^*)j - \int_0^j q^A_\phi(v,\beta^*)dv - t^A_\phi(0,\beta^*).$$

Hence,

$$\mathbb{E}_j t^A_\phi(j,\beta^*) = \int_0^J q^A_\phi(j,\beta^*)jf(j)dj - \int_0^J (\int_0^j q^A_\phi(j,\beta^*)dv)f(j)dj - t^A_\phi(0,\beta^*)$$
$$= \int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,\beta^*)\}f(j)dj - t^A_\phi(0,\beta^*),$$

where the last step follows from partial integration.

We thus replace $(BB)_k$ by:

$$\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j,\beta^*)\}f(j)dj - t^A_\phi(0,\beta^*) + t^P_\phi(k).$$

We denote the multiplier by δ_k.
Step 4: Lagrangian
We are ready to state the Lagrangian objective:

$$\max \mathcal{L} := \sum_{k=0}^{K} z_k^* U_\phi^P(k) + \rho[\int_0^J (q_\phi^A(v,\beta^*)dv + t_\phi^A(0,\beta^*) - U_D^A(J,\beta^*)]$$

$$+\sum_{k=0}^{\underline{K}-2} v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] + v_{\underline{K}-1,o}[U_\phi^P(\underline{K}-1,\underline{K}-1) - U_\phi^P(o,\underline{K}-1)]$$

$$+\sum_{k=0}^{K} \delta_k[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q_\phi^A(j,\beta^*)\}f(j)dj - t_\phi^A(0,\beta^*) + t_\phi^P(k)],$$

where $v_{k,k+1}$ and $v_{k-1,k}$ denote the multiplier on the upward adjacent (PIC) constraints.
Complementary slackness requires:

$$\rho[\int_0^J (q_\phi^A(v,\beta^*)dv + t_\phi^A(0,\beta^*) - U_D^A(J,\beta^*)] = 0,$$

$$v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] = 0 \ \ \forall k < \underline{K}-1,$$

$$\delta_k[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q_\phi^A(j,\beta^*)\}f(j)dj - t_\phi^A(0,\beta^*) + t_\phi^P(k)] = 0 \ \ \forall k.$$

Choice variables are the following:

$$\{t_\phi^A(0,\beta^*), t_\phi^P(k), q_\phi^A(\cdot,k), q_\phi^P(\cdot,k)\}_{k=0}^K.$$

Step 5: Necessary conditions for an optimum
We want to show that there exist $\{z_k^*\}_{k=0}^K$ and β^* such that $\phi = \phi^{RSW}$ solves the problem. That is, we treat z_k^* and β^* as choice variables. Throughout the presentation of the proof we assume that $\underline{K} < K$. If this condition fails, then $O = \emptyset$ and $v_{\underline{K}-1,k\in O} = v_{\underline{K}-1,\underline{K}} = 0$. We deal with the latter instance in footnotes.
As first step, we use the first order conditions of the choice variables $t_\phi^A(0,\beta^*)$ and $t_\phi^P(k)$ and construct (non-negative) multipliers being consistent with complementary slackness under the hypothesis that $\phi = \phi^{RSW}$ solves the problem.
The derivative with respect to $t_\phi^A(0,\beta^*)$ satisfies:

$$\rho - \sum_{k=0}^{K} \delta_k = 0 \tag{A.3}$$

The derivatives with respect to $t_\phi^P(k)$ satisfy:

$$- z_k^* - v_{k,k+1} + v_{k-1,k} = -\delta_k \ \ \forall k < \underline{K}-1, \tag{A.4}$$

$$- z_k^* - \sum_{o\in O} v_{k,o} + v_{k-1,k} = -\delta_k \ \ \text{for} \ \ k = \underline{K}-1, \tag{A.5}$$

$$- z_k^* + v_{\underline{K}-1,k} = -\delta_k \ \ \forall k \in O. \tag{A.6}$$

At this point it becomes handy to pool the types in O. We choose $z_k^* := z_o^* > 0$ and $\delta_k := 0$ for any $k \in O$ and define $\sum_{k=\underline{K}}^{K} z_k^* =: \alpha$. Summing (A.3) over k leads:

$$1 = \sum_{k=0}^{\underline{K}-1} \delta_k.$$

Iterating on (A.4)-(A.6) motivates the following choice of the multipliers:

$$v_{\underline{K}-1,k} := z_k^* - \delta_k = z_o^* \;\; \forall k \in O,$$

$$v_{0,1} := \beta_0^* - z_0^*,$$

$$v_{k-1,k} := \sum_{v=1}^{k-1} (\delta_v - z_v^*) + \alpha, \;\; \forall k \in \{2, ..., \underline{K} - 1\},$$

$$\alpha = \beta_0^* - z_0^*.$$

We choose $\beta_k^* = z_k^*$ for all $\underline{K} > k > 0$. For all $k < \underline{K}$ we choose $\delta_k = \beta_k^*$.[4]
These choices leave us with the following requirements on α, z_0^*, β^*, which are verified later:

$$(i) \;\; v_{0,1} = \beta_0^* - z_0^* \geq 0, \quad (ii) \;\; \sum_{k=0}^{K} z_k^* = z_0^* + \alpha + \sum_{k=1}^{\underline{K}-1} \beta_k^* = 1, \quad (iii) \;\; \alpha = \beta_0^* - z_0^* \quad \text{(F2)}$$

Step 6: Reformulation of the Lagrangian terms: $(PIC)^+$

We aim to manipulate the Lagrangian in order to derive an expression that allows us to solve the maximization problem point-wise. As first step, consider the following expression:

$$A := \sum_{k=0}^{K} U_\phi^P(k) + \sum_{k=0}^{\underline{K}-1} v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] + \sum_{k=0}^{\underline{K}-1} \delta_k [t_\phi^P(k)].$$

Note that

$$v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] = -v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k)) - (t_\phi^P(k+1) - t_\phi^P(k))].$$

Substituting in A and using the first order conditions with respect to $t_\phi^P(k)$, (A.4), leads:

$$A = \sum_{k=0}^{K} z_k^* \theta_k q_\phi^P(k) - \sum_{k=0}^{\underline{K}-1} v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k)) - (t_\phi^P(k+1) - t_\phi^P(k))] + \sum_{k=0}^{\underline{K}-1} (\delta_k - z_k^*)[t_\phi^P(k)]$$

$$= \sum_{k=0}^{K} z_k^* \theta_k q_\phi^P(k) - \sum_{k=0}^{\underline{K}-1} v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k))].$$

[4] If $\underline{K} > K$, we choose $\beta_k^* = z_k^*$ for all $K > k > 0$ and $\beta_k^* = \delta_k \;\; \forall k \in \Theta^P$. Moreover, we determine z_K^*, z_0^* and β_K^* (or δ_K),β_0^* later on via $1 - (z_K^* + z_0^*) = 1 - \sum_{k=1}^{K-1} \beta_k^* = 1 - (\beta_K + \beta_0)$. Define $z_K^* - \delta_K =: \alpha > 0$. Then $v_{K-1,K} := z_K^* - \beta_K^* = \alpha$, $v_{k-1,k} = \alpha$ for any $k > 1$ and $v_{0,1} = \beta_0^* - z_0^* \geq 0$.

Applying some algebra manipulations, we receive:

$$A = \sum_{k=1}^{\underline{K}-1} q_\phi^P(k)[\theta_k(z_k^* + v_{k,k+1}) - v_{k-1,k}\theta_{k-1}] + \sum_{k=\underline{K}}^{K} q_\phi^P(k)z_k^*\theta_k + (z_0^* + v_{0,1})q_\phi^P(0)\theta_0.$$

Again, by the first order conditions (A.4) and the assumed form of the multipliers:

$$q_\phi^P(k)[\theta_k(z_k^* + v_{k,k+1}) - v_{k-1,k}\theta_{k-1}] = q_\phi^P(k)[\delta_k\theta_k + (\theta_k - \theta_{k-1})v_{k-1,k}]$$

$$= q_\phi^P(k)[\delta_k\theta_k + (\theta_k - \theta_{k-1})(\sum_{v=1}^{k-1}(\delta_v - z_v^*) + \alpha)] = q_\phi^P(k)[\beta_k^*\theta_k + (\theta_k - \theta_{k-1})\alpha)],$$

whenever $0 < k < \underline{K}$.
A becomes:

$$A = \sum_{k=1}^{\underline{K}-1} q_\phi^P(k)[\beta_k^*\theta_k + (\theta_k - \theta_{k-1})\alpha)] + \sum_{k=\underline{K}}^{K} q_\phi^P(k)z_k^*\theta_k + (z_0^*)q_\phi^P(0)\theta_0.$$

Step 7: Reformulation of the Lagrangian, continued

We want to use the previous results to reformulate the Lagrangian. First, observe that $\mathcal{L} = A + B$ with B being defined as:

$$B := \rho[\int_0^J q_\phi^A(v, \beta^*)dv + U_\phi^A(0, \beta^*) - U_D^A(J, \beta^*)]$$

$$+(\sum_{k=0}^{K}\delta_k)[\int_0^J \{(j - \frac{1 - F(j)}{f(j)})q_\phi^A(j, \beta^*)\}f(j)dj - U_\phi^A(0, \beta^*)].$$

Making use of $\sum_{k=0}^{K}\delta_k = \rho = 1$, we can reformulate B:

$$B = \int_0^J \{(j + \frac{F(j)}{f(j)})q_\phi^A(j, \beta^*)\}f(j)dj - U_D^A(J, \beta^*).$$

$\mathcal{L}$ becomes:

$$\mathcal{L} = \sum_{k=1}^{\underline{K}-1} \int_0^J \{q_\phi^P(j,k)[\theta_k + (\theta_k - \theta_{k-1})\frac{\alpha}{\beta_k^*}] + (j + \frac{F(j)}{f(j)})q_\phi^A(j,k)\}\beta_k^* f(j)dj$$

$$-U_D^A(J, \beta^*) + \sum_{k=\underline{K}}^{K} q_\phi^P(k)z_k^*\theta_k + z_0^* q_\phi^P(0)\theta_0 + \int_0^J (j + \frac{F(j)}{f(j)})q_\phi^A(j,0)f(j)\beta_0 dj$$

$$+F(0)\sum_{k=1}^{\underline{K}-1} q_\phi^P(0,k)[\beta_k^*\theta_k + (\theta_k - \theta_{k-1})\alpha)].$$

Step 8: Construction of the variables by point-wise maximization

Let $q_{\phi^*}(\cdot)$ be the point-wise maximizer of $\mathcal{L}$. We choose the variables that are not determined up to now, $z_0^*, \{\beta_k^*\}_{k=0}^{\underline{K}-1}, \alpha$, such that $q_{\phi^*}(\cdot) = q_{RSW}(\cdot)$.
For any $k \in O$ it is straightforward to observe that $q_{\phi^*}^P(k) = 1$. Moreover, $\theta_0 = 0$ implies $q_{\phi^*}^P(j,0) = 0$ and $q_{\phi^*}^A(j,0) = 1\ \forall j \in \Theta^A$.

These both decisions are consistent with the RSW allotment policy.
Now consider $k \in \{1, ..., \underline{K}-1\}$. We want to show that we can find $z_0^*, \{\beta_k^*\}_{k=0}^{\underline{K}-1}, \alpha$ that satisfy feasibility and induce the same allotment policy as the RSW. In order to do so, we fix a generic k in the above interval and show that we can construct β_k^* such that the allotment policy is identical to the RSW. Afterwards, we verify that we have constructed probabilities. β_k^* is the solution to:

$$\theta_k + (\theta_k - \theta_{k-1})\frac{\alpha}{\beta_k^*} = j(k) + \frac{F(j(k))}{f(j(k))}.$$

Recall from the construction of the RSW that the right-hand side is equal to:

$$\theta_k + (\theta_k - \theta_{k-1})\frac{\sum_{v=k}^{K}(1-\delta_v^*)}{\delta_k^*}.$$

We thus equate:

$$\theta_k + (\theta_k - \theta_{k-1})\frac{\alpha}{\beta_k^*} = \theta_k + (\theta_k - \theta_{k-1})\frac{\sum_{v=k}^{K}(1-\delta_v^*)}{\delta_k^*}.$$

$$\iff \beta_k^* = \alpha\frac{\delta_k^*}{\sum\limits_{v=k}^{K}(1-\delta_v^*)}.$$

Note, since $\sum_{v=k}^{K}(1-\delta_v^*) > 0$ (see the proof of lemma 2.2) each β_k^* is weakly above zero and bounded. We thus need to show that by choosing α appropriately, we can assure $\sum_{k=1}^{\underline{K}-1}\beta_k^* + \beta_0^* = 1$. This implies that we can interpret each β_k^* as probability. Define $\Psi := \sum_{k=1}^{\underline{K}-1}\frac{\delta_k^*}{\sum_{v=k}^{K}(1-\delta_v^*)}$, and observe that: $\sum_{k=1}^{\underline{K}-1}\beta_k^* = \alpha\Psi$. Moreover, since $\sum_{k=1}^{K}\beta_k^* = \sum_{k=1}^{\underline{K}-1} z_k^*$ we need $z_0^* + \alpha = 1 - \alpha\Psi$ to assure that the welfare weights sum to 1 (i.e., (ii) in (F2)). Set $z_0^* = \alpha$, then $\alpha = \frac{1}{2+\Psi}$. With this choice the condition on the probability reads:

$$\sum_{k=1}^{\underline{K}-1}\beta_k^* + \beta_0^* = 1 \quad \iff \alpha\Psi + \beta_0^* = 1$$

Hence,

$$\beta_0^* = \frac{2}{2+\Psi} = 2\alpha < 1.$$

For fixed Ψ, part (i) and (iii) of (F2) are left to verify. That is, $\beta_0^* - z_0^* = \alpha$ and $\beta_0^* - z_0^* \geq 0$.
The last condition holds because $z_0^* = \alpha$. The first condition is satisfied by the choice of β_0^*.[5]
Step 9: Optimality of point-wise maximization

[5]If $\underline{K} > K$, then we face the conditions:
(i) $\alpha = z_K^* - \beta_K^* \geq 0$,
(ii) $\beta_0^* - z_0^* \geq 0$,
(iii) $z_K^* + z_0^* = \beta_0^* + \beta_K^*$,
(iv) $\sum_{k=1}^{K-1}\beta_k^* + \beta_0^* + \beta_K^* = 1$.

Consider the reformulated Lagrangian objective (under the above constructed probabilities, welfare weights and multipliers). By construction, we know that q_{RSW} maximizes L point-wise, and ϕ^{RSW} is such that complementary slackness is satisfied. Since q_{RSW} is monotonic, we thus know that the solution to the relaxed program $(P2)'$ satisfies the ignored global incentive constraints. Since ϕ^{RSW} is also feasible in problem $(P2)$, we know it is a solution to this latter program.

A.5 Proof of Proposition 2.2

Let us first define ϕ^e:

Ex-ante optimal Allocation.

$$(q^a_{\phi^e}(j,k), t^a_{\phi^e}(j,k)) = \begin{cases} (1, h^a_{\phi^e}(j) + \mathbb{E}_k[h^p_{\phi^e}(k)] - h^p_{\phi^e}(k)) & \text{if } j > k\,, \\ (0, h^a_{\phi^e}(j) + \mathbb{E}_k[h^p_{\phi^e}(k)] - h^p_{\phi^e}(k)) & \text{if } j \leq k. \end{cases}$$

$$(q^p_{\phi^e}(j,k), t^p_{\phi^e}(j,k)) = \begin{cases} (0, h^p_{\phi^e}(k) - \mathbb{E}_k[h^p_{\phi^e}(k)] - h^a_{\phi^e}(j)) & \text{if } j > k\,, \\ (1, h^p_{\phi^e}(k) - \mathbb{E}_k[h^p_{\phi^e}(k)] - h^a_{\phi^e}(j)) & \text{if } j \leq k. \end{cases}$$

Where

$$h^a_{\phi^e}(j) = q^a_{\phi^e}(j)j - \int_0^j q^a_{\phi^e}(v)dv,$$

$$h^p_{\phi^e}(k) = q^p_{\phi^e}(k)\theta_{k-1} - \sum_{v=1}^{k-1}(\theta_v - \theta_{v-1})q^p_{\phi^e}(v), \quad \text{if } \ k > 1$$

$$h^p_{\phi^e}(1) = h^p_{\phi^e}(0) = 0.$$

We prove that ϕ^e is an ex-ante optimal allocation by applying Lemma A.2, i.e., we specify $Z(k) = B(k) \ \ \forall k \in \Theta^P$ and verify that ϕ^e is a solution to $(P3)_z$:
By the definition of ϕ^e, it holds that $q^A_{\phi^e}(j) = q^A_D(j) \ \ \forall j \in \Theta^A$ and $q^P_{\phi^e}(k) = q^P_D(k) \ \ \forall k \in \Theta^P$. Moreover, $U^A_{\phi^e}(j) = U^A_D(j) \ \ \forall j \in \Theta^A$.
Given this, the measure that weights the different agent types individual rationality constraints, $\Lambda(j)$, being such that $\Lambda(j) = F(j) \ \ \forall j \in \Theta^A$ is consistent with condition (ii).
Condition (iv) of lemma A.2 now becomes:

$$q_{\phi^e} \in \arg\max_q \sum_{k=0}^{K} \int_0^J [q^P_\phi(j,k)\theta_k + jq^A_\phi(j,k)]\beta_k f(j)dj.$$

Condition (iv) is satisfied, since ϕ^e induces efficient collusion by definition. Condition (iii) has slack, since $Z(k) = B(k)$ (i.e., neither the principal's upward nor

---

Using the formula for β^*_k, i.e., $\beta^*_k = \alpha\frac{\delta^*_k}{\sum_{v=k}^{K}(1-\delta^*_v)}$, together with (i) we first observe: $\beta^*_K = \delta^*_K z^*_K$. Thus $\alpha = (1-\delta^*_K)z^*_K > 0$ satisfies (i). For given z^*_0, we choose $\beta^*_0 := \alpha + z^*_0$, and thus (iii) is satisfied. (ii) is satisfied, since $-z^*_0 + \beta^*_0 = -z^*_0 + \alpha + z^*_0 > 0$. Finally observe that (iv) is satisfied if $\alpha\Psi + \alpha + z^*_0 = 1$, that is, $\alpha = \frac{1-z^*_0}{1+\Psi}$, or $z^*_K = \frac{1-z^*_0}{(1-\delta^*_K)(1+\Psi)}$. It remains to show that we can find $1 > z^*_K, z^*_0 > 0$ satisfying this constraint. Note that the right-hand side is continuous and decreases in z^*_0. We set $z^*_0 = 1$ then $z^*_K = 0$. Hence, if we choose $z^*_0 < 1$ but close enough to 1, then a feasible z^*_K is determined by the above representation of constraint (iv).

downward incentive constraints bind). Note, the transfer scheme $\{t^P_{\phi^e}(k)\}_{k=0}^K$, defined above, satisfies the principal's upward adjacent incentive constraints with equality.

A.6 Proof of Lemma 2.7

An allocation that leads to efficient collusion cannot be supported in equilibrium if it leaves the largest principal type with strictly lower payoff than $U^P_{RSW}(K)$. Therefore, choose without loss of generality the most preferred ex-ante optimal allocation of principal type K, ϕ^e.

We want to verify that this allocation leaves every type of the principal with weakly larger payoff than the RSW allocation, if it leaves type K with weakly larger payoff than $U^P_{RSW}(K)$. By hypothesis:

$$U^P_{\phi^e}(K) \geq U^P_{RSW}(K).$$

Using the incentive-compatible representation of utilities, we thus have:

$$U^P_{\phi^e}(K) = q^P_{\phi^e}(K)(\theta_K - \theta_{K-1}) + U^P_{\phi^e}(K-1)$$

$$\geq q^P_{RSW}(K)(\theta_K - \theta_{K-1}) + U^P_{RSW}(K-1) = U^P_{RSW}(K).$$

Since $q^P_{\phi^e}(K) = q^P_{\phi^e}(K) \leq q^P_{RSW}(K)$, the above inequality implies:

$$U^P_{\phi^e}(K-1) \geq U^P_{RSW}(K-1).$$

Making use once again of the incentive-compatible representations of utilities:

$$U^P_{\phi^e}(K-1) = q^P_{\phi^e}(K-1)(\theta_{K-1} - \theta_{K-2}) + U^P_{\phi^e}(K-2)$$

$$\geq q^P_{RSW}(K-1)(\theta_{K-1} - \theta_{K-2}) + U^P_{RSW}(K-2) = U^P_{RSW}(K-1).$$

Since we know that $q^P_{RSW}(k) \geq q^P_{\phi^e}(k)$ for all $k \in \Theta^P$, by the fact that the RSW allocation implies inefficient collusion in favor of the principal, the above inequality implies that $U^P_{\phi^e}(K-2) \geq U^P_{RSW}(K-2)$. Continuing with this reasoning for all types below $K-2$, it follows that every type of principal prefers ϕ^e to the RSW allocation.

A.7 Proof of Lemma 2.8

Consider the following condition:

Condition 1. *Suppose that*
(i) $J = \Delta + \theta_K$, $\theta_k - \theta_{k-1} = \theta_{k+1} - \theta_k := \Delta$ *for all* $k \in \Theta^P$ *and there exist arbitrarily small, but positive,* δ_1 *such that :*

$$\max_{k \in \{1,\ldots,K\}} \left| \frac{(\theta_k - \theta_{k-1})}{\beta_k} B(k-1) - \frac{F(\theta_k)}{f(\theta_k)} \right| = \delta_1,$$

(ii) For any type profile with $j = \theta_k$, it is the case that

$$\frac{F(\theta_k)}{f(\theta_k)} \geq (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k},$$

(iii) $\frac{F(\theta_k)}{f(\theta_k)} - \frac{(\theta_k - \theta_{k-1})}{\beta_k} B(k-1)$ *is weakly increasing in k.*

(iv) $f(j)$ is differentiable for any $j > 0$.

First, note that distributions that δ-close distributions imply condition 1 if $\delta := \max_{k>0} \tilde{f}(k)$ is sufficiently small. This is the case, because the principal's type distribution, as being specified in definition 2.8, satisfies for any $k > 0$:

$$\frac{(\theta_k - \theta_{k-1})}{\beta_k} B(k-1) = \Delta \frac{F(\theta_k)}{\tilde{f}(k) + \Delta f(\theta_k)},$$

with $\tilde{f}(k)$ non-negative. Hence, (i) and (ii) are satisfied. Observe also, the difference defined in (iii) reads:

$$\frac{F(\theta_k)}{f(\theta_k)}\left(\frac{\frac{\tilde{f}(k)}{f}}{\frac{\tilde{f}(k)}{f(k)+\Delta}}\right)$$

Both terms are weakly increasing, and so is the product. Because $f(\cdot)$ is assumed to be differentiable, $\tilde{f}(k)$ becomes arbitrarily small if Δ becomes arbitrarily small. The maximal δ_1, or $\tilde{f}(k)$, will be jointly determined by equations (A.11) and (A.12), presented below.

Assume that condition 1 is satisfied. We show that $\phi_{z_K=1} = \phi^e$. To do so, let us solve $(\tilde{P3})_z$ with $z_K = 1$. We show that the solution, $\phi_{z_K=1}$, is equal to ϕ^e. Given this, it follows by lemma 2.7 that ϕ^e is an equilibrium allocation.

Let us claim that $\phi_{z_K=1} = \phi^e$. Given this, we know that $U_D^A(j) = U_{\phi^e}^A(j) \quad \forall j \in \Theta^A$ and thus can choose any measure $\Lambda(j)$ such that condition (iv) of lemma A.2 is satisfied, i.e.:

$$q_{\phi_{z_K=1}} \in \arg\max_q \sum_{k=0}^{K} \int_0^J [q_\phi^P(j,k)\{\theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k}\}$$

$$+ (j - \frac{\Lambda^-(j) - F(j)}{f(j)}) q_\phi^A(j,k)]\beta_k f(j) dj.$$

We construct $\Lambda(j)$ the following way:

For $j = \theta_k = 0$, let:

$$\Lambda(0) = 0.$$

For all $0 < j < J$, such that there exists some k with $\theta_k = j$, let:

$$\Lambda(j = \theta_k) = F(\theta_k) - \frac{f(\theta_k)}{\frac{\beta_k}{\theta_k - \theta_{k-1}}} B(k-1), \tag{A.7}$$

which is positive by condition 1 (iii).

Choose $\Lambda(J) = 1$.

For any $j \in (\theta_{k-1}, \theta_k)$ with $k \geq 1$ let

$$\Lambda(j) = (j - \theta_{k-1})\frac{\Lambda(\theta_k) - \Lambda(\theta_{k-1})}{\theta_k - \theta_{k-1}} + \Lambda(\theta_{k-1}),$$

and for $j \in (\theta_K, J)$ let

$$\Lambda(j) = (j - \theta_{k-1})\frac{\tilde{\Lambda}^{-}(J) - \Lambda(\theta_{k-1})}{\theta_k - \theta_{k-1}} + \Lambda(\theta_{k-1}),$$

where

$$\tilde{\Lambda}^{-}(J) := F(J) - \frac{f(J)}{\frac{\beta_K}{\theta_K - \theta_{K-1}}}B(K-1).$$

First, observe that by condition 1 (iii) and (iv), $\Lambda(j) \geq 0 \;\; \forall j$ and $\Lambda(j) - \Lambda(j') \geq 0$ for any $j \geq j'$. Therefore, condition (ii) of lemma A.2 is satisfied. Moreover, for future reference note that $F(j) \geq \Lambda(j)$ for all j. To see this, fix any $j \in (\theta_{k-1}, \theta_k)$ and observe:

$$F(j) - \Lambda(j) \geq 0$$

$$\iff F(j) - (j - \theta_{k-1})\frac{\Lambda(\theta_k) - \Lambda(\theta_{k-1})}{\theta_k - \theta_{k-1}} - \Lambda(\theta_{k-1}) \geq 0$$

By the definition of $\Lambda(\theta_{k-1})$, i.e., (A.7):

$$\iff F(j) - F(\theta_{k-1}) + B(k-2)\frac{f(\theta_{k-1})}{\beta_{k-1}}(\theta_k - \theta_{k-1}) \geq \frac{(j - \theta_{k-1})}{(\theta_k - \theta_{k-1})}(\Lambda(\theta_k) - \Lambda(\theta_{k-1}))$$

$$\iff \frac{F(j) - F(\theta_{k-1})}{j - \theta_{k-1}} + B(k-2)\frac{f(\theta_{k-1})}{\beta_{k-1}}\frac{(\theta_k - \theta_{k-1})}{j - \theta_{k-1}} \geq \frac{1}{(\theta_k - \theta_{k-1})}(\Lambda(\theta_k) - \Lambda(\theta_{k-1})) \tag{A.8}$$

By definition of $\Lambda(\theta_k)$, (A.7), the right-hand side of (A.8) reads:

$$\frac{F(\theta_k) - F(\theta_{k-1})}{\theta_k - \theta_{k-1}} - \frac{f(\theta_k)}{\beta_k}B(k-1) + \frac{f(\theta_{k-1})}{\beta_{k-1}}B(k-2)$$

Moreover, since the left-hand side of (A.8) satisfies

$$\frac{F(j) - F(\theta_{k-1})}{j - \theta_{k-1}} + B(k-2)\frac{f(\theta_{k-1})}{\beta_{k-1}}\frac{(\theta_k - \theta_{k-1})}{j - \theta_{k-1}} > \frac{F(j) - F(\theta_{k-1})}{\theta_k - \theta_{k-1}} + B(k-2)\frac{f(\theta_{k-1})}{\beta_{k-1}},$$

the implication that is to be shown holds if

$$\frac{f(\theta_k)}{\beta_k}B(k-1)(\theta_k - \theta_{k-1}) \geq F(\theta_k) - F(j).$$

$$\iff \frac{(\theta_k - \theta_{k-1})}{\beta_k}B(k-1) \geq \frac{F(\theta_k) - F(j)}{f(\theta_k)} \tag{A.9}$$

This is true because the right-hand side of (A.7) can be bounded strictly from above by $\frac{F(\theta_k)}{f(\theta_k)}$. This bound is, by condition 1 (ii), arbitrarily close to the left-hand side of (A.7).

It is left to show that (iv) of lemma A.2 is satisfied.

Denote the virtual valuations by

$$VV(j) := j - \frac{\Lambda^{-}(j) - F(j)}{f(j)},$$

.

$$VV(k) := \theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k}.$$

Optimality implies that, for any given type profile, the player with the higher virtual valuation receives the right. By construction of $\Lambda(j)$, it follows that $VV(j) = VV(k)$ whenever $j = \theta_k$. Thus, assuming the virtual valuations to be non-decreasing, we can allot the right to the principal whenever $\theta_k \geq j$, and to the agent whenever $j > \theta_k$. By condition 1, (i) and assumption 1 we know that the principal's virtual valuation is non-decreasing. It is therefore left to show that $VV(j)$ is non-decreasing. To do so, take the (left-hand side) derivative with respect to j. We denote by $\lambda^{-}(j)$ the left-hand side derivative of $\Lambda(j)$, i.e., $\lambda^{-}(j) := lim_{j'<j, j' \to j}|\frac{\Lambda^{-}(j)-\Lambda^{-}(j')}{j-j'}|$:

$$\frac{dVV(j)}{dj} = 2 - \frac{f'(j)}{f^2(j)}F(j) + \frac{f'(j)}{f^2(j)}\Lambda^{-}(j) - \frac{\lambda^{-}(j)}{f^2(j)}. \tag{A.10}$$

First, observe that

$$2 - \frac{f'(j)}{f^2(j)}F(j) + \frac{f'(j)}{f^2(j)}\Lambda^{-}(j) > 0. \tag{A.11}$$

To see this, note that $j + \frac{F(j)}{f(j)}$ is increasing by assumption 1. Hence: $2 - \frac{f'(j)}{f^2(j)}F(j) > 0$. Now, if $f'(j) \geq 0$, then inequality (A.11) is clearly satisfied. If $f'(j) < 0$, then the inequality (A.11) is also satisfied, since $\Lambda^{-}(j) \leq F(j)$ as shown above.

Secondly, observe that $\max_{j\in\Theta^A}|\lambda^{-}(j)| = \max_{k\in\Theta^P}|\frac{\Lambda(\theta_k)-\Lambda(\theta_{k-1})}{\theta_k-\theta_{k-1}}| = \max_{j\in\Theta^A}|\frac{F(\theta_k)}{\theta_k-\theta_{k-1}} - \frac{f(\theta_k)}{\beta_k}B(k-1) - \frac{F(\theta_{k-1})}{\theta_k-\theta_{k-1}} + \frac{f(\theta_{k-1})}{\beta_{k-1}}B(k-2)|$ is arbitrarily small, since condition 1 (ii) implies that $\forall k \in \Theta^P$ there exists a positive, but arbitrarily small δ such that

$$f(\theta_k) = \frac{F(\theta_k)}{\frac{(\theta_k-\theta_{k-1})}{\beta_k}B(k-1)} - \delta\frac{f(\theta_k)}{\frac{(\theta_k-\theta_{k-1})}{\beta_k}B(k-1)}. \tag{A.12}$$

Hence, the right-hand side of equation (A.10) is increasing in j.

A.8 Proof of Proposition 2.3

Step 1

Define $w(j) := \sum_{v=2}^{K}\mathbb{1}[j \in (\theta_{v-1}, \theta_v]]\theta_{v-1}$ and denote the c.p.f. of $w(\cdot)$ by $\tilde{F}(\cdot)$. Note that $\tilde{F}(\theta_k)$ =Prob$(j \leq \theta_{k+1}) = F(\theta_{k+1})$.

Observe that the linear-separable transfer scheme, h, associated with ϕ^e, being defined in the proof of proposition 2.2, is such that $\mathbb{E}_{(j,k)\in\Theta}[h^P_{\phi^e}(k) + h^A_{\phi^e}(j)] = \mathbb{E}_{(j,k)\in\Theta}[\min\{\theta_k, w(j)\}]$ and thus $U^P_{\phi^e}(0) = \mathbb{E}_{(j,k)\in\Theta}[\min\{\theta_k, w(j)\}]$ and $U^P_{\phi^e}(K) = \sum_{k=1}^{K}F(\theta_k)(\theta_k - \theta_{k-1}) + \mathbb{E}_{(j,k)\in\Theta}[\min\{\theta_k, w(j)\}]$.

Define $V := \min\{\theta_k, w(j)\}$, and note that the c.p.f. reads $Pr(V \leq v) = \tilde{F}(v) + B(v) - \tilde{F}(v)B(v)$, with support on Θ^P. As $Pr(v)$ is a right-continuous function, we can rewrite the expectation operator as Stieltjes integral. Hence,

$$\mathbb{E}_{(j,k)\in\Theta}[\min\{w(j),\theta_k\}] = \int_0^{\theta_K} v dPr(v)$$

For any $v > 0$ let $Pr^-(v) := \lim_{v'\to v, v'<v} Pr(v')$, i.e., the left-hand side limit and define $Pr^-(0) := Pr(0)$. Define $B^-(\cdot)$ and $F^-(\cdot)$ similarly. Thus, $Pr^-(\cdot) = \tilde{F}^-(\cdot) + B^-(\cdot) - \tilde{F}^-(\cdot)B^-(\cdot)$. Applying partial integration leads:

$$\mathbb{E}_{(j,k)\in\Theta}[\min\{w(j),\theta_k\}] = \theta_K - \int_0^{\theta_K} Pr^-(v)dv.$$

Hence,

$$U^P_{\phi^e}(K) = \sum_{k=1}^{K} F(\theta_k)(\theta_k - \theta_{k-1}) + \theta_K - \int_0^{\theta_K} Pr^-(v)dv$$

$$\iff U^P_{\phi^e}(K) = \sum_{k=1}^{K} F(\theta_k)(\theta_k - \theta_{k-1}) + \theta_K - \int_0^{\theta_K} B^-(v)(1-\tilde{F}^-(v))dv - \int_0^{\theta_K} \tilde{F}^-(v)dv,$$

$$= \theta_K - \int_0^{\theta_K} B^-(v)(1 - \tilde{F}^-(v))dv$$

Step 2

Recall that the RSW payoff reads

$$U^P_{RSW}(K) = (\theta_K - \theta_{\underline{K}-1}) + U^P_{RSW}(\underline{K} - 1) = \sum_{v=1}^{K} F(j(v))(\theta_v - \theta_{v-1}) < \theta_K,$$

where the last inequality follows from assumption 2.

Step 3

We now want to verify the existence of $\tilde{B}(\cdot)$ such that the highest principal type is indifferent between ϕ^e and the RSW allocation.

Observe, if $\beta_0 = B(0) \to 1$, then

$$U^P_{\phi^e}(K) \to \sum_{k=1}^{K} F(\theta_k)(\theta_k - \theta_{k-1})$$

Hence,

$$U^P_{RSW}(K) - U^P_{\phi^e}(K) \to \sum_{k=1}^{K} F(j(k))(\theta_k - \theta_{k-1}) - \sum_{k=1}^{K} F(\theta_k)(\theta_k - \theta_{k-1}) > 0,$$

because $F(j(k)) > F(\theta_k)$ for all $k > 0$. Moreover, if $\beta_K \to 1$, i.e., $B(k)$ converges to 0 for any $k < K$, then $U^P_{\phi^e}(K) \to \theta_K$ and thus

$$U^P_{RSW}(K) - U^P_{\phi^e}(K) \to \sum_{k=1}^{K} F(j(k))(\theta_k - \theta_{k-1}) - \theta_K < 0.$$

Hence, efficient collusion can be supported.

Finally, note that $U^P_{RSW}(K) - U^P_{\phi^e}(K)$ is continuous in every element of the vector β.

It therefore follows by the intermediate value theorem that there exists some $\tilde{B}(\cdot)$ such that the above difference is equal to 0. Of course, this $\tilde{B}(\cdot)$ can be chosen to have full-support on Θ^P.

Step 4

Finally, we show that if $\hat{B}(\cdot)$ first order stochastically dominates $\check{B}(\cdot)$, then the principal's payoff from efficient collusion when $B(\cdot) = \hat{B}(\cdot)$ is larger than if $B(\cdot) = \check{B}(\cdot)$. The respective payoff difference reads:

$$U^P_{\phi^e}(K)|_{B(\cdot)=\hat{B}(\cdot)} - U^P_{\phi^e}(K)|_{B(\cdot)=\check{B}(\cdot)}$$

$$= -\int_0^J \hat{B}^-(v)(1-\tilde{F}^-(v))dv - (-\int_0^J \check{B}(v)(1-\tilde{F}^-(v))dv)$$

$$= \int_0^J (\check{B}^-(v) - \hat{B}^-(v))(1-\tilde{F}^-(v))dv.$$

This term is positive as $\check{B}(v) \geq \hat{B}(v) \quad \forall v \in \Theta^P$.

A.9 Lemma A.2: Characterization of undominated (equilibria) Allocations

We first give a non-constructive characterization of undominated (equilibria) allocations. This characterization is extremely useful to prove proposition 2.2, theorem 2.11 and lemma 2.8. Moreover, lemma 2.6 is a direct consequence of this characterization.

Definition A.1. *An allocation ϕ is β-undominated, or undominated, if there exist weakly positive welfare weights $\{z_k\}_{k=0}^K$ (with $z_k \geq 0 \ \ \forall k \in \Theta^P$) such that ϕ solves*

$$\max_{\phi} \sum_{k=0}^{K} z_k U^P_\phi(k) \qquad (\tilde{P}4)_z$$

subject to:

$$\phi \text{ is } \tilde{\beta}\text{−feasible}$$

Whenever ϕ is a solution to $(P4)_z$ being defined similar to problem $(\tilde{P}4)_z$ but augmented by the constraints

$$(PIR)^S_k \quad U^P_\phi(k) \geq U^P_{RSW}(k) \quad \forall k \in \Theta^P,$$

if of collusion in a SPA, or by the constraints

$$(PIR)^F_k \quad U^P_\phi(k) \geq \theta_k \quad \forall k \in \Theta^P,$$

if of collusion in a FPA, we say that ϕ is an undominated equilibrium allocation.

Note that the definition of a $\tilde{\beta}$-feasible allocation is different for cases of collusion in a FPA or SPA.

Lemma A.2. *For given welfare weights, fix a β-feasible allocation, ϕ, satisfying (PIR). Suppose that*

(i) there exists a number $\tilde{z} \geq 0$ together with a (weakly increasing) function, $\tilde{Z}(k)$, such that $\tilde{Z}(0) = \tilde{z}_0$, $\tilde{Z}(K) = \tilde{z}$ and $\tilde{Z}(k) - \tilde{Z}(k-1) = \tilde{z}_k$, where $\tilde{z}_k > 0$ only if $U^P_\phi(k) =$

$U^P_S(k)$, and $U^P_S(k)$ denotes, depending whether we consider collusion in the SPA or FPA, either $U^P_{RSW}(k)$ or θ_k,
(ii) there exists a measure, $\Lambda(j)$, such that : $d\Lambda(j) := \lambda(j) > 0$ only if $U^A_\phi(j) = U^A_D(j)$, $\Lambda(j)$ is weakly increasing, $\Lambda(0) \in [0,1]$ and $\Lambda(J) = 1$. Let $\Lambda^-(j) := lim_{j'\to j, j'<j}\Lambda(j')$ for any $j > 0$ and $\Lambda^-(0) := \lambda(0)$,
(iii) $[U^P_\phi(k+1,k+1) - U^P_\phi(k,k+1)](Z(k) + \tilde{Z}(k) - (1+\tilde{z})B(k))^+ = 0$,
$[U^P_\phi(k-1,k-1) - U^P_\phi(k,k-1)]((1+\tilde{z})B(k-1) - Z(k-1) - \tilde{Z}(k-1))^+ = 0$.
Moreover, $q^P_\phi(k) \geq q^P_\phi(k-1)$, and for any $k > 0$ there is $\tilde{b}_k \geq 0$ such that $\tilde{b}_k[q^P_\phi(k) - q^P_\phi(k-1)] = 0$.
(iv) For any $(j,k) \in \Theta$ $q_\phi(j,k)$ satisfies:

$$q_\phi(j,k) \in \arg\max_{q^A,q^P\in[0,1]^2, q^A+q^P\leq 1}[q^P\{\theta_k(1+\tilde{z}) + \frac{\tilde{b}_k - \tilde{b}_{k+1}}{\beta_k})$$
$$-(\theta_{k+1}-\theta_k)\frac{(Z(k)+\tilde{Z}(k)-(1+\tilde{z})B(k))^+}{\beta_k} + (\theta_k-\theta_{k-1})\frac{((1+\tilde{z})B(k-1)-Z(k-1)-\tilde{Z}(k-1))^+}{\beta_k}\}$$
$$+(1+\tilde{z})(j - \frac{\Lambda^-(j)-F(j)}{f(j)})q^A],$$

and the resulting $q^A_\phi(j)$ is weakly increasing in j.
Then ϕ is a solution to $(\tilde{P}4)_z$. Moreover, if we drop condition (i), then ϕ is a solution to $(\tilde{P}4)_z$ if it satisfies (ii),(iii) and (iv) with $\tilde{z} = 0$ and $\tilde{Z}(k) = 0$ $\forall k \in \Theta^P$.

For completeness, note that lemma A.2 provides sufficient conditions for optimality. By this, environments, that is, $F(\cdot), B(\cdot)$ and $Z(\cdot)$ that give rise to bunching allotments with respect to the agent are ignored. We do not consider bunching allotments, because those situations are not relevant for any result presented in this paper. To derive a complete characterization, optimal control techniques as in Jullien, 2000 could be employed.

Proof. Define $(P4)'_z$ by weakening the ex-post budget-balanced constraint in $(\tilde{P}4)_z$ to ex-ante budget-balanced $(\tilde{BB})$. A standard result from mechanism design implies that in programs as the one considered here, ex-post balanced-budget can be weakened to ex-ante budget-balanced (e.g., see Börgers and Norman (2009)).
We are given some specified $\{z_k\}_{k=0}^K$. Suppose ϕ satisfies the hypothesis of lemma A.2. We want to show that it is solution to $(P4)'_z$.
We start by relaxing $(P4)'_z$.
Step 1: (AIC)
We work with the necessary conditions and later verify that the sufficient conditions for incentive compatibility are satisfied.
Note, $\frac{\partial U^A_\phi(\hat{j},j)}{\partial j} \leq 1$, for all $q^A_\phi(\hat{j}), t^A_\phi(\hat{j})$.
By Theorem 2 of Milgrom and Segal, 2002 we thus know that the incentive-compatible form of the utility is given by:

$$U^A_\phi(j) = \int_0^j q^A_\phi(v)dv + U^A_\phi(0),$$

where $U^A_\phi(0) = q^A_\phi(0)0 + t^A_\phi(0)$ is the utility of the lowest agent type and $t^A_\phi(0)$.

Moreover, $U^A_\phi(j)$ is absolutely continuous in j.

Step 2: (AIR)

By the above necessary condition, individual rationality reads :

$$\int_0^j q^A_\phi(v)dv + U^A_\phi(0) - U^A_D(j) \geq 0.$$

Define

$$\tilde{J}_\phi := arg \min_{j\in\Theta^A}\{\int_0^j q^A_\phi(v)dv - U^A_D(j)\},$$

which might be a set of (possibly infinitely many) types.

At this point, take any $\tilde{j} \in \tilde{J}_{\phi^*}$. It immediately follows that $U^A_\phi(\tilde{j}) \geq U^A_D(\tilde{j})$ implies $(AIR)^j$ for all j.

We replace (AIR) by $U^A_\phi(\tilde{j}) - U^A_D(\tilde{j}) \geq 0$ and denote the relevant multiplier by ρ.

Step 3: $(\tilde{BB})$

Making use of step 1, we obtain the following representation for the transfers:

$$t^A_\phi(j) = q^A_\phi(j)j - \int_0^j q^A_\phi(v)dv - t^A_\phi(0) \quad \forall j \in \Theta^A.$$

Applying partial integration leads:

$$\mathbb{E}_j t^A_\phi(j) = \int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j)\}f(j)dj - t^A_\phi(0).$$

We thus replace $(\tilde{BB})$ by:

$$\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j)\}f(j)dj - t^A_\phi(0) + \mathbb{E}_k t^P_\phi(k).$$

We denote the multiplier by δ.

Step 4: Lagrangian

We are ready to state the Lagrangian objective:

$$\max \mathcal{L} := \sum_{k=0}^K z_k U^P_\phi(k)$$

$$+\rho[\int_0^{\tilde{j}} (q^A_\phi(v)dv + U^A_\phi(0) - U^A_D(\tilde{j})]$$

$$+\sum_{k=0}^{K-1} v_{k,k+1}[U^P_\phi(k,k) - U^P_\phi(k+1,k)] + \sum_{k=1}^{K} v_{k,k-1}[U^P_\phi(k,k) - U^P_\phi(k-1,k)]$$

$$+\sum_{k=0}^{K} \tilde{z}_k[U^P_\phi(k) - U^P_S(k)]$$

$$+\delta\sum_{k=0}^{K} \beta_k[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q^A_\phi(j)\}f(j)dj - t^A_\phi(0) + t^P_\phi(k)]$$

$$+\sum_{k=1}^{K}\tilde{b}_k[q_\phi^P(k)-q_\phi^P(k-1)],$$

where $v_{k,k+1}$ and $v_{k-1,k}$ denote the multiplier on the adjacent (PIC) constraints with $v_{-1,0}=v_{K,K+1}=0$. Let $\tilde{z}_k$ be the multiplier on $(PIR)_k$. $\tilde{b}_k$ is the multiplier on the condition that any allocation that satisfies (PIC) must feature a weakly increasing allotment policy with respect to the principal.
Complementary slackness requires:

$$\rho[\int_0^{\tilde{j}} q_\phi^A(v)dv+U_\phi^A(0)-U_D^A(\tilde{j})]=0,$$

$$v_{k,k+1}[U_\phi^P(k,k)-U_\phi^P(k+1,k)]=0\ \ \forall k<K,$$

$$v_{k,k-1}[U_\phi^P(k,k)-U_\phi^P(k-1,k)]=0\ \ \forall k>0,$$

$$\tilde{z}_k[U_\phi^P(k)-U_S^P(k)]=0\ \ \forall k,$$

$$\delta[\int_0^J\{(j-\frac{1-F(j)}{f(j)})q_\phi^A(j)\}f(j)dj-t_\phi^A(0,\tilde{\beta})+\mathbb{E}_k t_\phi^P(k)]=0\ \ \forall k,$$

$$\tilde{b}_k[q_\phi^P(k)-q_\phi^P(k-1)]=0\ \ \forall k. \tag{A.13}$$

The choice variables are the following:

$$\{t_\phi^A(0),t_\phi^P(k),q_\phi^A(\cdot,k),q_\phi^P(\cdot,k)\}_{k=0}^K.$$

Step 6: Necessary conditions for an optimum

As first step we use the first order conditions of the choice variables $t_\phi^A(0)$ and $t_\phi^P(\Theta^P)$ and construct (non-negative) multipliers being consistent with complementary slackness under the hypothesis that ϕ solves the problem.
The derivative with respect to $t_\phi^A(0)$ satisfies:

$$\rho-\delta\sum_{k=0}^{K}\beta_k=0.$$

Hence $\rho=\delta$.
For any k, the derivative with respect to $t_\phi^P(k)$ satisfies :

$$-z_k-\tilde{z}_k-v_{k,k+1}+v_{k+1,k}-v_{k,k-1}+v_{k-1,k}=-\beta_k\delta. \tag{A.14}$$

Summing over the K conditions (A.14) leads:

$$1+\tilde{z}=\delta\sum_{k=0}^{K}\beta_k,$$

with $\tilde{z}:=\sum_{k=0}^{K}\tilde{z}_k$.
We thus set

$$\rho:=1+\tilde{z},\delta:=(1+\tilde{z}).$$

Iterating on the conditions (A.14) suggests the following choice of $v_{\cdot,\cdot}$:

$$v_{k-1,k} - v_{k,k-1} = (1+\tilde{z})B(k-1) - Z(k-1) - \tilde{Z}(k-1),$$

where $Z(k) := \sum_{v=0}^{k} z_v$, $\tilde{Z}(k) := \sum_{v=0}^{k} \tilde{z}_v$.
By complementary slackness $v_{k-1,k}v_{k,k-1} = 0$.[6] Thus set:

$$v_{k-1,k} := ((1+\tilde{z})P(k-1) - Z(k-1) - \tilde{Z}(k-1))^+, v_{k,k-1} := (Z(k-1) + \tilde{Z}(k-1) - (1+\tilde{z})B(k-1))^+.$$

Step 7: Reformulation of the Lagrangian terms: (PIC)
Given the above necessary conditions, we aim to manipulate the Lagrangian in order to derive an expression that allows us to verify the optimality of ϕ. As first step, consider the following expression:

$$A := \sum_{k=0}^{K}(z_k+\tilde{z}_k)U_\phi^P(k) + \sum_{k=0}^{K-1} v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] + \sum_{k=1}^{K} v_{k,k-1}[U_\phi^P(k,k) - U_\phi^P(k-1,k)]$$
$$+ \sum_{k=0}^{K} \delta\beta_k[t_\phi^P(k)].$$

Note that

$$v_{k,k+1}[U_\phi^P(k,k) - U_\phi^P(k+1,k)] = -v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k)) - (t_\phi^P(k+1) - t_\phi^P(k))],$$

and

$$v_{k,k-1}[U_\phi^P(k,k) - U_\phi^P(k-1,k)] = v_{k-1,k}[\theta_k(q_\phi^P(k) - q_\phi^P(k-1)) - (t_\phi^P(k) - t_\phi^P(k-1))].$$

Substituting in A and using (A.14) leads:

$$A = \sum_{k=0}^{K}(z_k+\tilde{z}_k)\theta_k q_\phi^P(k) - \sum_{k=0}^{K-1} v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k))] + \sum_{k=1}^{K} v_{k,k-1}[\theta_k(q_\phi^P(k) - q_\phi^P(k-1))].$$

Adding (A.13) we receive

$$A = \sum_{k=0}^{K}[(z_k+\tilde{z}_k)\theta_k + \tilde{b}_k - \tilde{b}_{k+1}))]q_\phi^P(k) - \sum_{k=0}^{K-1} v_{k,k+1}[\theta_k(q_\phi^P(k+1) - q_\phi^P(k))] + \sum_{k=1}^{K} v_{k,k-1}[\theta_k(q_\phi^P(k) - q_\phi^P(k-1))]$$

where $\tilde{b}_0 = \tilde{b}_{-1} := 0$. Applying some algebra manipulations we receive:

$$A = \sum_{k=0}^{K} q_\phi^P(k)[\tilde{b}_k - \tilde{b}_{k+1} + \theta_k(z_k + \tilde{z}_k + v_{k,k+1} + v_{k,k-1}) - \theta_{k+1}v_{k+1,k} - v_{k-1,k}\theta_{k-1}].$$

Again, by the first order conditions w.r.t. $t^P(k)$ (A.14) and the assumed form of the multipliers:

$$\theta_k(z_k+\tilde{z}_k+v_{k,k+1}+v_{k,k-1}) - \theta_{k+1}v_{k+1,k} - v_{k-1,k}\theta_{k-1} = \theta_k\beta_k(1+\tilde{z}) - (\theta_{k+1}-\theta_k)v_{k+1,k} + (\theta_k - \theta_{k-1})v_{k-1,k}$$

[6] Suppose $v_{k-1,k}, v_{k,k-1} > 0$. In this instance complementary slackness requires $U_\phi^P(k-1,k-1) = U_\phi^P(k,k-1)$ and $U_\phi^P(k,k) = U_\phi^P(k,k-1)$, which implies $U_\phi^P(k-1,k-1) = U_\phi^P(k,k)$. A contradiction, whenever $q_\phi^P(k) > 0$ (which is satisfied for any $k > 0$).

$= \theta_k\beta_k(1+\tilde{z})-(\theta_{k+1}-\theta_k)(Z(k)+\tilde{Z}(k)-(1+\tilde{z})B(k))^{+}+(\theta_k-\theta_{k-1})((1+\tilde{z})B(k-1)-Z(k-1)-\tilde{Z}(k-1))^{+}.$

A now reads:

$$A = \sum_{k=0}^{K} q_\phi^P(k)\beta_k\{\theta_k(1+\tilde{z}) + \frac{\tilde{b}_k - \tilde{b}_{k+1}}{\beta_k})$$

$$-(\theta_{k+1}-\theta_k)\frac{(Z(k)+\tilde{Z}(k)-(1+\tilde{z})B(k))^{+}}{\beta_k}+(\theta_k-\theta_{k-1})\frac{((1+\tilde{z})B(k-1)-Z(k-1)-\tilde{Z}(k-1))^{+}}{\beta_k}\}.$$

Step 7: Reformulation of the Lagrangian, continued

We want to use the previous results to reformulate the Lagrangian. First observe that $\mathcal{L} = A + B$, with B being defined as:

$$B := \rho[\int_0^{\tilde{j}} q_\phi^A(v)dv + U_\phi^A(0) - U_D^A(\tilde{j})]$$

$$+\delta[\int_0^J \{(j - \frac{1-F(j)}{f(j)})q_\phi^A(j)\}f(j)dj - U_\phi^A(0)].$$

If we are able to show that ϕ maximizes $\mathcal{L}$ (and complementary slackness is satisfied), then we know that ϕ is a solution to $(\tilde{P}3)_z$.

We first manipulate

$$\rho[\int_0^{\tilde{j}} q_\phi^A(v)dv + U_\phi^A(0) - U_D^A(\tilde{j})]$$

further. Reconsider $\tilde{J}_\phi$. By hypothesis, there exists $\Lambda(j) := \int^j d\Lambda(j)$ such that $\Lambda(J) = 1$, $\Lambda(0) = 0$ and $d\Lambda(j) := \lambda(j) > 0$ only if $j \in \tilde{J}_\phi$. We perform the following manipulations:

$$\rho[\int_0^{\tilde{j}} q_\phi^A(v)dv + U_\phi^A(0) - U_D^A(\tilde{j})]$$

$$= \rho\Lambda(J)[\int_0^{\tilde{j}} q_\phi^A(v)dv + U_\phi^A(0) - U_D^A(\tilde{j})]$$

$$= \rho\int_0^J [\int_0^{\tilde{j}} q_\phi^A(v)dv - U_D^A(\tilde{j}) + U_\phi^A(0)]d\Lambda(j).$$

Since $\lambda(j) > 0$ only if $j \in \tilde{J}_\phi$ the above term is equal to

$$\rho\int_0^J [\int_0^{j} q_\phi^A(v)dv - U_D^A(j) + U_\phi^A(0)]d\Lambda(j)$$

$$= \rho\{\int_0^J [\int_0^{j} q_\phi^A(v)dv]d\Lambda(j) - \int_0^J U_D^A(j)d\Lambda(j) + U_\phi^A(0)\}.$$

Applying partial integration leads:

$$\rho\{\Lambda(J)\int_0^J dU_\phi^A(j) - \int_0^J \Lambda^-(j)dU_\phi^A(j) - \int_0^J U_D^A(j)d\Lambda(j)) + U_\phi^A(0)]\}$$

$$= \rho\{\int_0^J (1-\Lambda^-(j))dU_\phi^A(j) - \int_0^J U_D^A(j)d\Lambda(j)) + U_\phi^A(0)]\}$$

$$= \rho\{[\int_0^J q_\phi^A(j))(1-\Lambda^-(j))dj - \int_0^J U_D^A(j)d\Lambda(j) + U_\phi^A(0)]\},$$

where the last step follows from the absolute continuity of $U_\phi^A(j)$.
Making, in addition, use of $\delta = \rho = (1+\tilde{z})$ we can reformulate B:

$$B = \int_0^J \{(1+\tilde{z})(j - \frac{\Lambda^-(j) - F(j)}{f(j)})q_\phi^A(j)\}f(j)dj$$

$$-(1+\tilde{z})\int_0^J U_D^A(j)d\Lambda(j).$$

$\mathcal{L}$ reads:

$$\mathcal{L} = \sum_{k=0}^{K}\int_0^J [q_\phi^P(j,k)\{\theta_k(1+\tilde{z}+\frac{\tilde{b}_k - \tilde{b}_{k+1}}{\beta_k})$$

$$-(\theta_{k+1}-\theta_k)\frac{(Z(k)+\tilde{Z}(k)-(1+\tilde{z})B(k))^+}{\beta_k} + (\theta_k - \theta_{k-1})\frac{((1+\tilde{z})B(k-1)-Z(k-1)-\tilde{Z}(k-1))^+}{\beta_k}\}$$

$$+(1+\tilde{z})(j - \frac{\Lambda^-(j)-F(j)}{f(j)})q_\phi^A(j,k)]p_k f(j)dj$$

$$-(1+\tilde{z})\int_0^J U_D^A(j)d\Lambda(j) - \sum_{k=0}^{K}\tilde{z}_k U_S^P(k)$$

$$+F(0)\sum_{k=0}^{K} q_\phi^P(k,0)\beta_k\{\theta_k(1+\tilde{z}) + \tilde{b}_k - \tilde{b}_{k-1})$$

$$-(\theta_{k+1}-\theta_k)\frac{(Z(k)+\tilde{Z}(k)-(1+\tilde{z})B(k))^+}{\beta_k} + (\theta_k - \theta_{k-1})\frac{((1+\tilde{z})B(k-1)-Z(k-1)-\tilde{Z}(k-1))^+}{\beta_k}\}.$$

Step 8

By assumption, ϕ maximizes the Lagrangian point-wise and satisfies complementary slackness. Moreover, q_ϕ is monotonic by assumption. Therefore, ϕ also solves the original program $(\tilde{P4})_z$. □

A.10 Proof of Lemma 2.9

Let $b^P(k)$ ($b^A(j)$) be principal type k's (agent type j's) bid function. The following strategies constitute an equilibrium of the FPA, played with belief $\tilde{\beta}$ about the principal, and a degenerate belief concentrated on the largest agent type J:
Define $\hat{k} := \arg\max_{\check{k}\in\Theta^P} \tilde{B}(\check{k})(J-\theta_{\check{k}})$ and for any k such that $k \geq \hat{k}+1$, define $b_k := \frac{\tilde{B}(\hat{k})}{\tilde{B}(k)}\theta_{\hat{k}} + \frac{\sum_{v=\hat{k}+1}^{k}\tilde{\beta}_v}{\tilde{B}(k)}J$ and $b_{\hat{k}} := \theta_{\hat{k}}$. Moreover, for any k, such that $k \geq \hat{k}+1$, define $\alpha_k := \prod_{v=k}^{K}\frac{\theta_k - b_k}{\theta_k - b_{k-1}}$.

- Principal type $k \leq \hat{k}$ submits a bid equal to $b^P(k) = \theta_k$.

- Principal type k with $k \geq \hat{k}+1$ submits a bid b on the interval $(b_{k-1}, b_k]$ with cdf

$$F_k^P(b) = \frac{\tilde{B}(k-1)}{\tilde{\beta}_k}\frac{b-b_{k-1}}{J-b}.$$

- Agent type J submits a bid on the interval $(b_{\hat{k}} = \theta_{\hat{k}}, b_K]$ according to the cdf:

$$F^A(b) = \mathbb{1}[b \in (b_{k-1}, b_k]]\{\alpha_k \frac{b - b_{k-1}}{\theta_k - b} + F^A(b_{k-1})\},$$

and point mass at $\theta_{\hat{k}}$ equals $F^A(\theta_{\hat{k}}) = 1 - \sum_{k=\hat{k}+1}^{K} \alpha_k \frac{b_k - b_{k-1}}{\theta_k - b_k}$.

- Agent type j's with $j < J$ optimal bid, $b^A(j)$, is given by

$$b^A(j) \in \arg\max_{\check{k} \leq \hat{k}} \tilde{B}(\check{k})(j - \theta_{\check{k}}).$$

Proof. Let us define by $\Pi^A(J, b)$ the payoff of agent type J from playing the FPA given the above stated strategies of the principal. Similarly, define by $\Pi^P(k, b)$ the payoff of principal type k.

Observe first, given the principal's strategy, and the fact, which is verified later on, that agent type J is indifferent between submitting any bid on the interval $[\theta_{\hat{k}}, b_K]$ it must be the case that any agent type $j < J$ optimally bids $b^A(j) \leq \theta_{\hat{k}}$. Hence, any such type solves

$$b^A(j) \in \arg\max_{\check{k} \leq \hat{k}} \tilde{B}(\check{k})(j - \theta_{\check{k}}).$$

Moreover, given the principal's degenerate belief concentrated on the largest agent type J, and the latter's strategy, every principal type, k with $k \leq \hat{k}$ maximizes his expected payoff when bidding $b^P(k) \leq \theta_k$. Hence, $b^P(k) = \theta_k$ is an optimal action. Moreover, also observe that

$$\theta_k - b_k \geq 0 \quad \iff (\theta_k - J) + \frac{\tilde{B}(\hat{k})}{\tilde{B}(k)}(J - \theta_{\hat{k}}) \geq 0.$$

This is true, because by definition $\hat{k} := \arg\max_{\check{k} \in \Theta^P} \tilde{B}(\check{k})(J - \theta_{\check{k}})$, that is,

$$\tilde{B}(\hat{k})(J - \theta_{\hat{k}}) \geq \tilde{B}(k)(J - \theta_k) \quad \forall k \in \Theta^P.$$

Also, note that $b_k \geq 0$ for all $k \geq \hat{k}$, because $J\tilde{B}(k) \geq (J - \theta_{\hat{k}})\tilde{B}(\hat{k})$.

Now, let us verify that the principal's bid cdf's are indeed probability measures. $F_k^P(b_{k-1}) = 0$ is satisfied by construction. To see that $F_k^P(b_k) = \frac{\tilde{B}(k-1)}{\tilde{\beta}_k} \frac{b_k - b_{k-1}}{J - b_k} = 1$ is true, note:

$$J - b_k = \frac{\tilde{B}(\hat{k})}{\tilde{B}(k)}(J - \theta_{\hat{k}})$$

and

$$\begin{aligned} b_k - b_{k-1} &= (\frac{\tilde{B}(\hat{k})}{\tilde{B}(k)} - \frac{\tilde{B}(\hat{k})}{\tilde{B}(k-1)})\theta_{\hat{k}} + (\frac{\sum_{v=\hat{k}+1}^{k} \tilde{\beta}_v}{\tilde{B}(k)} - \frac{\sum_{v=\hat{k}+1}^{k-1} \tilde{\beta}_v}{\tilde{B}(k-1)})J \\ &= (\frac{-\tilde{\beta}_k \tilde{B}(\hat{k})}{\tilde{B}(k)\tilde{B}(k-1)})\theta_{\hat{k}} + (\frac{\tilde{\beta}_k \tilde{B}(\hat{k})}{\tilde{B}(k)\tilde{B}(k-1)})J. \end{aligned}$$

Hence,

$$\frac{\tilde{B}(k-1)}{\tilde{\beta}_k}(b_k - b_{k-1}) = \frac{\tilde{B}(\hat{k})}{\tilde{B}(k)}(J - \theta_{\hat{k}}).$$

Therefore,

$$F^P(b_k) = \frac{\tilde{B}(\hat{k})}{\tilde{B}(k)} \frac{\tilde{B}(k)}{\tilde{B}(\hat{k})} = 1.$$

Now consider agent type J, and let us verify that he is indifferent between bidding on the interval $[\theta_{\hat{k}}, b_K]$. To do so, note that when submitting a bid $b \in [b_{k-1}, b_k]$, agent type J wins according to the probability $\tilde{\beta}_k F_k^P(b) + \tilde{B}(k-1)$. Any such bid implies the payoff, $\Pi^A(J, b)$, with

$$\Pi^A(J, b \in (b_{k-1}, b_k]) = (J - b)[\tilde{\beta}_k F_k^P(b) + \tilde{B}(k-1)]$$

Observe that for any $b \in (b_{k-1}, b_k]$ it is the case that

$$\frac{d\Pi(J,b)}{db}|_{b\in(b_{k-1},b_k]} = \tilde{\beta}_k \frac{\tilde{B}(k-1)}{\tilde{\beta}_k} - \tilde{B}(k-1) = 0$$

Moreover, since the principal's bid strategy has no point mass, agent type J is indifferent between any bid on $[b_{\hat{k}}, b_K]$.
To argue for global optimality, note that

$$\Pi^A(J, b = \theta_{\hat{k}}) = \tilde{B}(\hat{k})(J - \theta_{\hat{k}})$$

and

$$\Pi^A(J, , b = b_K) = J - b_K = \tilde{B}(\hat{k})(J - \theta_{\hat{k}})$$

Submitting a bid above b_K cannot be optimal, as a bid of b_K is sufficient to win the auction with certainty. Moreover, by the definition of $\hat{k}$ it follows that agent type J does not want to submit a bid strictly below $\theta_{\hat{k}}$.
Now, consider a principal type k with $k \geq \hat{k} + 1$. Let us verify that he is indifferent between bidding $b \in (b_{k-1}, b_k]$. To do so, note that when submitting a bid $b \in (b_{k-1}, b_k]$, principal type k wins the auction with probability $\alpha_k \frac{b - b_{k-1}}{\theta_k - b} + F^A(b_{k-1})$ where

$$\begin{aligned} F^A(b_{k-1}) &= \sum_{v=\hat{k}+1}^{k-1} \alpha_v \frac{b_v - b_{v-1}}{\theta_v - b_v} + 1 - \sum_{v=\hat{k}+1}^{K} \alpha_v \frac{b_v - b_{v-1}}{\theta_v - b_v} \\ &= 1 - \sum_{v=k+1}^{K} \alpha_v \frac{b_v - b_{v-1}}{\theta_v - b_v} - \alpha_k \frac{b_k - b_{k-1}}{\theta_k - b_k}. \end{aligned}$$

Any such bid implies the payoff, $\Pi^P(k, b)$, :

$$\Pi^P(k, b \in (b_{k-1}, b_k]) = (\theta_k - b)[\alpha_k \frac{b - b_{k-1}}{\theta_k - b} + F^A(b_{k-1})]$$

Hence ,

$$\frac{d\Pi^P(k,b)}{db}|_{b\in(b_{k-1},b_k])} = \alpha_k - F^A(b_{k-1}) = \alpha_k \frac{\theta_k - b_{k-1}}{\theta_k - b_k} + \sum_{v=k+1}^{K} \alpha_v \frac{b_v - b_{v-1}}{\theta_v - b_v} - 1$$

We want to argue that this quantity is zero and principal type k is thus indifferent between submitting any bid on the specified interval. To do so, let us start with

$k = K$.[7] $\frac{d\Pi^P(K,b)}{db}|_{b\in(b_{K-1},b_K])}$ reads:

$$\alpha_K \frac{\theta_K - b_{K-1}}{\theta_K - b_K} - 1 = 0 \quad \Longleftrightarrow \alpha_K = \frac{\theta_K - b_K}{\theta_K - b_{K-1}},$$

which is true by the definition of α_K. For any $k \in \{\hat{k}+1, ..., K-1\}$ we want to verify the indifference condition by backward induction. Let us fix $K-1$. We need:

$$\frac{d\Pi^P(K-1,b)}{db}|_{b\in(b_{K-2},b_{K-1}])} = 0 \Longleftrightarrow \quad \alpha_{K-1}\frac{\theta_{K-1} - b_{K-2}}{\theta_{K-1} - b_{K-1}} + \alpha_K \frac{b_K - b_{K-1}}{\theta_K - b_K} = 1$$

$$\Longleftrightarrow \alpha_{K-1} = (1 - \frac{b_K - b_{K-1}}{\theta_K - b_{K-1}})\frac{\theta_{K-1} - b_{K-1}}{\theta_{K-1} - b_{K-2}} = \frac{\theta_K - b_K}{\theta_K - b_{K-1}}\frac{\theta_{K-1} - b_{K-1}}{\theta_{K-1} - b_{K-2}},$$

which follows from the definition of α_{K-1}. Now, the induction hypothesis is

$$\alpha_k - F^A(b_{k-1}) = 0, \quad \forall k \in \{\hat{k}+2, ..., K-2\} \tag{A.15}$$

Consider $k = \hat{k}+1$. We need to verify that

$$\alpha_{\hat{k}+1} - F^A(b_{\hat{k}}) = 0$$

Let us use

$$F^A(b_{\hat{k}}) = F^A(b_{\hat{k}+1}) - \alpha_{\hat{k}+1}\frac{b_{\hat{k}+1} - b_{\hat{k}}}{\theta_{\hat{k}+1} - b_{\hat{k}+1}},$$

to derive at:

$$\alpha_{\hat{k}+1} - F^A(b_{\hat{k}}) = \alpha_{\hat{k}+1} - F^A(b_{\hat{k}+1}) + \alpha_{\hat{k}+1}\frac{b_{\hat{k}+1} - b_{\hat{k}}}{\theta_{\hat{k}+1} - b_{\hat{k}+1}}$$

$$= \alpha_{\hat{k}+1}\frac{\theta_{\hat{k}+1} - b_{\hat{k}}}{\theta_{\hat{k}+1} - b_{\hat{k}+1}} - F^A(b_{\hat{k}+1})$$

Making use of the induction hypothesis, (A.15), this reads:

$$= \alpha_{\hat{k}+1}\frac{\theta_{\hat{k}+1} - b_{\hat{k}}}{\theta_{\hat{k}+1} - b_{\hat{k}+1}} - \alpha_{\hat{k}+2} = \alpha_{\hat{k}+2}(\frac{\theta_{\hat{k}+1} - b_{\hat{k}+1}}{\theta_{\hat{k}+1} - b_{\hat{k}}}\frac{\theta_{\hat{k}+1} - b_{\hat{k}}}{\theta_{\hat{k}+1} - b_{\hat{k}+1}} - 1) = 0,$$

where the last step follows from the definition of α_k.
Let us argue for global optimality of the principal's strategy. We start with type $\hat{k}+1$. When submitting a bid weakly below $\theta_{\hat{k}}$, he receives payoff zero, which is lower than

$$F^A(b_{\hat{k}+1})(\theta_{\hat{k}+1} - b_{\hat{k}+1}) = \Pi(\hat{k}+1, b = b_{\hat{k}+1}).$$

Moreover, $\Pi^P(\hat{k}+1, b)$ is continuous on the interval $b \in \cup_{k\geq\hat{k}+1}(b_k, b_{k+1}]$, since there are no other point masses in agent type J's bidding strategy. $\Pi^P(\hat{k}+1, b)$ is decreasing on this interval, because for each sub-interval $(b_k, b_{k+1}]$, there is k with $k > \hat{k}+1$ being indifferent between any bid on this sub-interval. From a similar argument global optimality of $b^P(k)$ for any k with $k > \hat{k}+1$ follows.
Finally, observe that the stated strategy of the agent, $F^A(\cdot)$, is indeed a distribution.

[7] Recall that terms involving types that do not exist are zero, e.g., $\sum_{k+1}^K \alpha_v \frac{b_v - b_{v-1}}{\theta_v - b_v} = 0$

We have seen that for any $k \in \{\hat{k}, ..., K-1\}$ it holds that $F^A(b_k) = \alpha_{k+1}$. Using the definition of $F^A(b_K)$ and α_K, we thus observe: $F^A(b_K) = \alpha_K \frac{b_K - b_{K-1}}{\theta_K - b_K} + \alpha_K = 1$. Moreover, $F^A(\theta_{\hat{k}}) = F^A(b_{\hat{k}}) = \alpha_{\hat{k}+1} \geq 0$.

□

A.11 Proof of Lemma 2.10

Observe first, given any agent type j and any action, $d = \check{k}$, of the latter, it is a (weakly) dominant strategy for the principal to report his type truthfully: The principal receives payoff

$$\mathbb{1}[\hat{k} > \check{k}]\theta_k + \mathbb{1}[\hat{k} \leq \check{k}]\theta_{\check{k}},$$

from reporting to be $\hat{k}$. Hence, if $d = \check{k} \geq k$, any report $\hat{k} \leq \check{k}$, in particular $\hat{k} = k$, leaves principal type k with the highest possible payoff, given d. If $d = \check{k} < k$, the principal strictly prefers any report $\hat{k} > \check{k}$, in particular $\hat{k} = k$, to any other report.

Now consider the agent: We want to argue that any agent type j optimally submits report $d = \arg\max_{\check{k} \in \Theta^P} \tilde{B}(\check{k})(j - \theta_{\check{k}})$. By the construction of $\tilde{m}$, agent type j receives (ex-post) payoff equal to $\mathbb{1}[k \leq d][j - \theta_d]$, conditional that principal type k realizes and truthfully reports his type. Hence, for any belief $\tilde{\beta}$ the agent holds about the principal, the former optimally submits a report that maximizes $\sum_{k=0}^{K} \tilde{\beta}_k \mathbb{1}[k \leq d][j - \theta_d] = \tilde{B}(d)(j - \theta_d)$. Thus, by lemma 2.9 every agent type receives exactly his outside option.

A.12 Proof of Proposition 2.4

The on-path play of the equilibrium is such that every type of the principal offers the game form $\tilde{m}$, and the agent accepts the collusive side game.

If the principal offers the game form $\tilde{m}$ and the agent rejects, the FPA is played non-cooperatively with off-path belief about agent type J.

If the principal offers a game form m' that differs from $\tilde{m}$, the continuation game is played with a degenerate off-path belief concentrated on principal type 0, i.e., $\tilde{\beta}_0 = 1$.

Consider any continuation equilibrium starting at the beginning of period 2 after the proposal of m' with $m' \neq \tilde{m}$ and the agent holding a degenerate off-path belief concentrated on principal type 0. It is the agent's turn to decide whether to accept or to reject m'.

Suppose first the agent rejects m'. Let us consider the non-cooperative play of the FPA, being played with $\tilde{\beta}_0 = 1$. For any belief the principal might hold about the agent, the non-cooperative play of the FPA is such that principal type 0 bids his valuation, $\theta_0 = 0$, and agent type j bids θ_0 and wins the auction auction against principal type 0. Given this, every principal type above θ_0 optimally submits a bid above, but arbitrarily close to, θ_0, wins the auction and receives payoff $\theta_k - \theta_0$.

The (continuation) equilibrium play of the continuation game that starts with the agent's decision whether or whether not to accept m' is such that the agent ratifies m' only if he expects to receive larger payoff than his expected outside option, the payoff from playing the auction non-cooperatively against principal type 0, i.e., $U_D^A(j, \tilde{\beta}_0 = 1) = \tilde{\beta}_0(j - \theta_0) = j$.

Hence, any $(\tilde{\beta}_0 = 1)$-feasible allocation leaves the agent with payoff weakly larger than $U_D^A(j, \tilde{\beta}_0 = 1) = jq_D^A(j, 0) - \theta_0 = j$.
As last step, we argue that principal type's k payoff cannot be larger than θ_k. We start with type 0. He cannot receive larger payoff than θ_0, since this is the solution of the following problem, which defines types 0 largest payoff in a world where his type were commonly known:

$$\max_{\phi} U_\phi^P(0)$$

such that

$$\phi \quad \text{is} \quad \tilde{\beta}_0 = 1 - \text{feasible}$$

This is true, because a $(\tilde{\beta}_0 = 1)$-feasible allocation satisfies $\forall j \in \Theta^A$:

$$(AIR)_0^j \quad j - \theta_0 \leq U_\phi^A(j, k = 0),$$

$$(AIC)_0^j \quad j \in \arg\max_{\hat{j} \in \Theta^A} U_\phi^A(j, \hat{j}, k = 0),$$

$$(PIC)_0 \quad 0 \in \arg\max_{\hat{k} \in \Theta^P} U_\phi^P(k = 0, \hat{k}),$$

and

$$\forall (j, k) \in \Theta \quad (BB)_{j,k}.$$

If the allotment of the right is efficient, that is, agent type j with $j \geq \theta_0 = 0$ receives the right against principal type 0, $(AIR)_0^j$ is satisfied with equality for all types of the agent. Consequently, principal type 0 can absorb the entire collusive surplus. Moreover, since an efficient allocation is induced, the collusive surplus is also maximized.
Now consider principal type $k \geq 1$. An upper bound for the deviation-payoff he receives from the proposal of m' is given by his most preferred $(\tilde{\beta}_0 = 1)$-feasible allocation. This allocation is defined as solution of the following problem:

$$\max_{\phi} U_\phi^P(k)$$

such that $\forall j \in \Theta^A$

$$(AIR)_0^j \quad j - \theta_0 \leq U_\phi^A(j, k = 0),$$

$$(AIC)_0^j \quad j \in \arg\max_{\hat{j} \in \Theta^A} U_\phi^A(j, \hat{j}, k = 0),$$

$$(PIC)_0 \quad 0 \in \arg\max_{\hat{k} \in \Theta^P} U_\phi^P(k = 0, \hat{k}),$$

and

$$\forall (j, k) \in \Theta \quad (BB)_{j,k}.$$

Observe first that by $(PIC)_0$ it must be the case that $t_\phi^P(k) \geq 0$: If $t_\phi^P(k) < 0$ principal type 0 would report $\hat{k} > 0$ and receive payoff strictly larger than θ_0.
Given this, the solution of the above problem features $t_\phi^P(k) = 0$ and $q_\phi^P(\cdot, k) = 1$ if $k \geq 1$. As a consequence, principal type k receives at most the payoff θ_k, which is weakly lower than the payoff from offering $\tilde{m}$.

A.13 Proof of Theorem 2.1: lemmatas 2.11 and 2.12

Proof of Lemma 2.11

Step 1

To avoid pathological cases, let us verify that $\tilde{m}$ can be perturbed slightly such that every agent type has strict incentives to accept the side mechanism and the play of the perturbed version of $\tilde{m}$ specifies a unique equilibrium.

Consider the following simultaneous move mechanism, m^o. Let the agent's message / action be $d \in \Theta^P$. Let the principal's message /action be $\hat{k} \in \Theta^P$. Given the report profile $(d, \hat{k}) = (\check{k}, k)$, the following allotment and transfers result, where $\mu \in (0, 1)$, is a given parameter:

$\forall d \in \Theta^P$:

$$(q^A_{m^o}(d,k), t^A_{m^o}(d,k)) = \left\{ \begin{array}{ll} (0, -t^P_{m^o}(d,k)) & \forall k \in \Theta^P. \end{array}\right.$$

If $k = 0$:

$$(q^P_{m^o}(d,k), t^P_{m^o}(d,k)) = \left\{ \begin{array}{ll} (0, 2J) & \forall d \in \Theta^P \end{array}\right. .$$

$\forall k > 0:$

$$(q^P_{m^o}(d,k), t^P_{m^o}(d,k)) = \left\{ \begin{array}{ll} (\mu B(k), -U^P_{m^o}(k-1,k) + \frac{\theta_k + \theta_{k-1}}{2}\mu B(k)) & \forall d \in \Theta^P \end{array}\right. .$$

Observe that for every type report/action profile and any $\tilde{\beta}$, m^o induces a $\tilde{\beta}$-feasible allocation. Moreover, according to m^o every agent type receives strictly larger payoff than from playing the auction non-cooperatively against the principal. In addition, according to m^o the principal has strict incentives to report his type truthfully, for any action d, the agent might choose.

As a consequence, the mechanism $\hat{\tilde{m}}$, defined as convex combination between $\tilde{m}$ and m^o, is such that every principal type k has a unique optimal type report, $\hat{k} = k$, and the agent strictly prefers the play of $\tilde{m}$ to the play of the default game.

By proposing $\hat{\tilde{m}}$ with sufficiently high probability mass on $\tilde{m}$ principal type K receives payoff arbitrarily close to θ_K. Hence, only allocations which leave principal type K with payoff θ_K can arise in equilibrium.

Step 2

We next verify that the play of $\tilde{m}$ given any prior belief β implements the most preferred allocation of the largest principal type. Because of the quasilinear environment, there is a unique most preferred allocation. Since this allocation leaves type K with exactly the payoff θ_K, this in turn implies that no other allocation induces this payoff to the largest type of the principal. To do so, we apply lemma A.2 and drop condition (i). We choose $z_K = 1$ and $z_k = 0$ for any $k < K$.

Observe first that by lemma 2.9 every agent type's outside option is satisfied with equality. We therefore can choose $\Lambda(j) = F(j)\ \forall j \in \Theta^A$.

Moreover, for any $j \in \Theta^A$ define $\check{k}(j) := \max\{\arg\max_{\check{k} \in \Theta^P}(\check{k})(j - \theta_{\check{k}})\}$. Then it follows from the lemma 2.10 that the allocation implemented by $\tilde{m}$ with prior belief abut the principal, say $\phi_{\tilde{m}}$, reads:

$$(q^A_{\phi_{\tilde{m}}}(j,k), t^A_{\phi_{\tilde{m}}}(j,k)) = \left\{ \begin{array}{ll} (1, \theta_{\check{k}(j)}) & \text{if } k \leq \check{k}(j), \\ (0,0) & \text{if } k > \check{k}(j). \end{array}\right.$$

$$(q^P_{\phi_{\tilde{m}}}(j,k), t^P_{\phi_{\tilde{m}}}(j,k)) = \begin{cases} (0, -\theta_{\check{k}(j)}) & \text{if } \check{k}(j) \geq k, \\ (1, 0) & \text{if } \check{k}(j) < k. \end{cases}$$

Since this allocation is implemented by $\tilde{m}$, it is β-feasible. It is therefore left to verify that the above allocation is also the most preferred β-feasible allocation of the largest principal type, i.e., we need to verify that the allotment policy is consistent with condition (iv) of lemma A.2, that is, with point-wise maximization of the Lagrangian objective.

To do so, first consider the set $\check{\Theta}^P$ of principal types, such that $k \in \check{\Theta}^P$ if and only if $\exists j : \check{k}(j) = k$. Because $B(\check{k})(j - \theta_{\check{k}})$ is continuously increasing in j, for any $k > 0$ with $k \in \check{\Theta}^P$, there exists a threshold agent type $j_{\tilde{m}}(k)$, such that, $\check{k}(j_{\tilde{m}}(k)) = \{k', k\}$, where k' is k's downward-adjacent neighbor in $\check{\Theta}^P$. Hence,

$$B(k)(j_{\tilde{m}}(k) - \theta_k) = B(k')(j_{\tilde{m}}(k) - \theta_{k'})$$

$$\Rightarrow \quad j_{\tilde{m}}(k) = \theta_k + \frac{B(k')}{B(k) - B(k')}(\theta_k - \theta_{k'}) \quad \forall k \in \check{\Theta}^P \setminus 0. \tag{A.16}$$

Furthermore, because $B(\check{k})(j - \theta_{\check{k}})$ is increasing in j, $j_{\tilde{m}}(k)$ is non-decreasing in $k \in \check{\Theta}^P$.

Moreover, $j_{\tilde{m}}(\check{k}(J)) \leq J$ and

$$B(\check{k}(J))(J - \theta_{\check{k}(J)}) \geq B(k)(J - \theta_k) \quad \forall k > \check{k}(J)$$

$$\Rightarrow \quad J \leq \theta_k + (\theta_k - \theta_{\check{k}(J)})\frac{B(\check{k}(J))}{B(k) - B(\check{k}(J))} \quad \forall k > \check{k}(J).$$

Since the lowest agent type is zero and $\theta_0 = 0$, it follows that the lowest element in $\check{\Theta}^P$ is zero. Moreover, for any such $k \in \check{\Theta}^P$, there is a threshold agent type, $j_{\tilde{m}}(k)$ such that principal type k receives the right according to the play of $\tilde{m}$ if $j < j_{\tilde{m}}(k)$ and the agent receives the right against principal type k if $j \geq j_{\tilde{m}}(k)$. Now consider $k \notin \check{\Theta}^P$. Let k', k'' be two adjacent types in $\check{\Theta}^P$ and $k' < k < k''$. Then

$$B(k'')(j_{\tilde{m}}(k'') - \theta_{k''}) = B(k')(j_{\tilde{m}}(k'') - \theta_{k'}) > B(k)(j_{\tilde{m}}(k'') - \theta_k), \tag{A.17}$$

and thus k receives the right in every event in which type k'' receives the right.

Hence, the resulting allotment policy can equivalently be described by a sequence of increasing threshold agent types $\{j_{\tilde{m}}(k)\}_{k \in \Theta^P}$ according to:

$$q^A_{\phi_{\tilde{m}}}(j,k) = \begin{cases} 1 & \text{if j} \geq j_{\tilde{m}}(k), \\ 0 & \text{if } j < j_{\tilde{m}}(k). \end{cases}$$

$$q^P_{\phi_{\tilde{m}}}(j,k) = \begin{cases} 0 & \text{if } j_{\tilde{m}}(k) > \text{j}, \\ 1 & \text{if } j_{\tilde{m}}(k) \leq j. \end{cases}$$

In addition, observe that the allocation of the play of $\tilde{m}$ is such that the principal's upward adjacent incentive constraints are satisfied with equality, because for any

$k \in \Theta^P$ the following holds:

$$U^P_{\phi_{\tilde{m}}}(k+1,k) = F(j_{\tilde{m}}(k+1))\theta_k + \sum_{v=k+2}^{K} (F(j_{\tilde{m}}(v)) - F(j_{\tilde{m}}(v-1)))\theta_v$$

$$= F(j_{\tilde{m}}(k))\theta_k + \sum_{v=k+1}^{K} [F(j_{\tilde{m}}(v)) - F(j_{\tilde{m}}(v-1))]\theta_v) = U^P_{\phi_{\tilde{m}}}(k,k)$$

As a consequence, to verify condition (iv) of lemma A.2 we can define the increments of the Lagrangian objective as:

$$VV^A(j) := j - \frac{\Lambda(j) - F(j)}{f(j)} = j,$$

$$VV^P(k) := \theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k}, \quad \text{if } k-1, k \in \check{\Theta}^P.$$

Observe that for any $k-1, k \in \check{\Theta}^P$, it is the case that $VV^P(k) = VV^A(j(k))$. Because the agent's virtual valuation, $VV^A(j) = j$, is increasing in j, this also implies that the above described allotment policy is consistent with point-wise maximization of the Lagrangian objective, i.e., condition (iv) of lemma A.2 is satisfied, for any $\{k, j\} \in \check{\Theta}^P \times \Theta^A$. Moreover, for any $k \in \check{\Theta}^P$, $VV^P(k)$ is also weakly increasing, because $j_{\tilde{m}}(k)$ is, and for any $k > \check{k}(J)$ it holds that $VV^P(k) > J$.

It is left to consider those types of the principal $k \notin \check{\Theta}^P$. Let k', k'' be two adjacent types in $\check{\Theta}^P$ with $k' < k < k''$. According to the claimed optimal solution, k is pooled with type k''. Define the virtual valuation of type k the follwing way

$$VV^p(k) := \theta_k + (\theta_k - \theta_{k'})\frac{B(k')}{B(k) - B(k')} > j_{\tilde{m}}(k''), \tag{A.18}$$

where the inequality follows from (A.17).

To verify that this is indeed optimal, we make use of condition (iii) and pool the principal's virtual valuations the following way: Let

$$\tilde{b}_{k'+2} = (VV^P(k'+1) - j_{\tilde{m}}(k''))\beta_{k'+1},$$

and for any $k \in \{k'+2, ..., k''-1\}$ (if such a type exists), let

$$\tilde{b}_{k+1} = \sum_{v=k'+1}^{k} (VV^P(v) - j_{\tilde{m}}(k''))\beta_v.$$

By construction the resulting (pooled) virtual valuation, say $\tilde{VV}^P(k)$, being defined as

$$\tilde{VV}^P(k) := VV^P(k) + \frac{\tilde{b}_k}{\beta_k} - \frac{\tilde{b}_{k+1}}{\beta_k},$$

satisfies

$$\tilde{VV}^P(k) = j_{\tilde{m}}(k'').$$

Moreover, for any such (pooled) k it is the case that $\tilde{VV}^P(k) = j_{\tilde{m}}(k'') > j_{\tilde{m}}(k') = VV^P(k')$ and thus the above described allotment policy is consistent with condition

(iv) of lemma A.2.
We finally must verify that any such $\tilde{b}_{k+1}$ is non-negative. This follows by the fact that $k \notin \check{\Theta}^P$: Suppose to the contrary that for some $\tilde{b}_{k+1} < 0$, and take the lowest such $k > k'$. Then

$$\sum_{v=k'+1}^{k} (VV^P(v) - j_{\tilde{m}}(k''))\beta_v < 0$$

$$\Rightarrow \sum_{v=k'+1}^{k} [\theta_v B(v) - \theta_{v-1} B(v-1)] < (B(k) - B(k'))j_{\tilde{m}}(k'')$$

$$\Rightarrow B(k)\theta_k - B(k')\theta_{k'} < (B(k) - B(k'))j_{\tilde{m}}(k'')$$

$$\Rightarrow \theta_k + (\theta_k - \theta_{k'})\frac{B(k')}{B(k) - B(k')} < j_{\tilde{m}}(k'').$$

Hence,

$$B(k)(j_{\tilde{m}}(k'') - \theta_k) > B(k')(j_{\tilde{m}}(k'') - \theta_{k'}).$$

This is a contradiction to (A.18).

Proof of Lemma 2.12

Let both the principal and the agent hold fixed, but arbitrary beliefs about each other. Denote the corresponding cdfs by $\tilde{F}(\cdot)$ and $\tilde{B}(\cdot)$ and define by $\Pi^A(j)$ agent type j's expected payoff induced by the un-cooperative Bayes-Nash-equilibrium play of the FPA, in undominated bidding strategies. Let $\underline{j} := \inf\{j \in \Theta^A | \tilde{F}(j) > 0\}$. We first verify that $\Pi^A(\underline{j}) = \max_{\check{k} \in \Theta^P}\{\tilde{B}(\check{k})(\underline{j} - \theta_{\check{k}}), 0\}$. This establishes the uniqueness of the continuation equilibrium of the game that begins after the proposal of $\tilde{m}_\epsilon$. We then use this feature to prove that the grand game has a unique perfect Bayesian equilibrium.

Lemma A.3.

$$\Pi^A(\underline{j}) = \max_{\check{k} \in \Theta^P}\{\tilde{B}(\check{k})(\underline{j} - \theta_{\check{k}}), 0\}$$

Proof. First note that the agent can secure himself the payoff of $\max_{\check{k} \in \Theta^P}\{\tilde{B}(\check{k})j - \theta_{\check{k}}, 0\}$, by bidding 0 or $\check{\theta}_k$.
Therefore, assume without loss of generality that $\Pi^A(\underline{j}) > 0$. We show that there exists k, such that $\Pi^A(\underline{j}) = \tilde{B}(k)(j - \theta_k)$.
In the following we denote by $b^A(j)$ and $b^P(k)$ the agent's and the principal's equilibrium strategies, which might be bid distributions.

Observation A.1. *Take any $b \in supp(b^A(j))$ and any $j' > j$. Then, $b \leq \max supp(b^A(j'))$*

Proof. Suppose to the contrary that $b > \max supp(b^A(j'))$. Then, j' submits all his bids below j. Moreover, since $b^A(j)$ is an equilibrium strategy, the expected winning probability of j strictly increases in his bid, for all bids in the support of $b^A(j)$. Since both j and j' have the same belief about the principal, this implies that the expected winning probability of j' is lower than the one of j. A contradiction to incentive compatibility. □

Definition A.2. *For fixed equilibrium bidding strategy (or bid distribution) of the agent, denote by* $\Pi^P(j,k)$ *the expected payoff of principal type* k *induced by his equilibrium strategy, conditional on the realization of* j.

Observation A.2. *Let* $k' \in \{k \in supp(\tilde{\beta}) | \Pi^P(k') > 0\}$. *Then* $\Pi^P(k', \underline{j}) > 0$.

Proof. We want to show that $\Pi^P(\underline{j}, k) > 0$. Suppose that $\Pi^P(\underline{j}, k) = 0$. Then, there is $j > \underline{j}$ such that $\Pi^P(j,k) > 0$ and thus $\Pi^P(j,k) > \Pi^P(\underline{j}, \bar{k})$. Hence, there is a bid $b \in supp(b^P(k))$, $b < \theta_k$, such that submitting this bid implies that k wins with strictly positive probability against j. By observation A.1, this implies that he also wins with strictly positive probability against $\underline{j}$, when submitting this bid. Thus, $\Pi^P(\underline{j}, k) > 0$.

□

Observation A.3. *Suppose an agent and a principal type receive strictly positive payoffs. If the supports of their bid distributions are not disjoint, then not both bid distributions can have a mass point at the same bid.*

Proof. Any bid in the joint bid support is strictly lower than both the type of the agent and the principal. If both distributions have a mass point at the same bid, then at least one type has a better strategy: the strategy that has a mass point arbitrary above the mass point of the opponent. □

Now, take the lowest principal type that receives strictly positive payoff, say k. Observe that $\Pi^P(k, \underline{j}) > 0$ by observation A.2. Moreover, k's bid distribution cannot have a mass point: observation A.3 implies that he then receives zero payoff against $\underline{j}$. Moreover, $\min supp(b^A(\underline{j})) \geq \theta_{k-1}$, because otherwise θ_{k-1} could garner positive payoff, by bidding above, but close to $\min supp(b^A(\underline{j}))$. If $\min supp(b^A(\underline{j})) > \theta_{k-1}$, then k must have a mass point at $\min supp(b^A(\underline{j}))$. Otherwise $\underline{j}$ strictly prefers to bid θ_{k-1} than $supp(b^A(\underline{j}))$. Hence $\min supp(b^A(\underline{j})) = \theta_{k-1}$.
Because θ_{k-1} is in the support of $b^A(\underline{j})$, the payoff of the agent reads $\tilde{B}(k-1)(\underline{j} - \theta_{k-1})$.

□

Lemma A.4. *The unique continuation equilibrium that begins after the proposal of* $\tilde{m}_\epsilon$ *is such that all agent types accept the mechanism.*

Proof. Suppose that there exists a continuation equilibrium being such that some agent types reject the mechanism with positive probability. Denote the set of these agent types by Θ^A_R. The principal's belief about the agent can only have support in Θ^A_R. Moreover, for any $j \in \Theta^A_R$ it must be the case that $\Pi^A(j) \geq \max_{k \in \Theta^P}\{\tilde{B}(\check{k})(j - \theta_{\check{k}}), 0\} + \epsilon$. However, this is a contradiction to lemma A.3. □

Take any allocation, induced by some perfect Bayesian equilibrium. Call it ϕ and observe that ϕ is β-feasible, by inscrutability. By assumption $\phi \neq \phi^\beta_{\tilde{m}}$ and no principal type wants to propose $\tilde{m}$. Recall from lemma A.4 that K can secure himself payoff equals $\theta_K - \epsilon$ by proposing $\tilde{m}_\epsilon$. Hence, for any $\epsilon > 0$, it is the case that $U^P_\phi(K) \geq \theta_K - \epsilon$.
As in step 2 in the proof of lemma 2.11 we focus on the problem of finding principal type K's payoff maximizing allocation, out of the class of all β-feasible allocations.

Recall, for any prior β, the unique solution to this problem is $\phi^{\beta}_{\tilde{m}}$. Since ϕ is an equilibrium allocation, it is β-feasible and therefore $U^P_{\phi}(K) < \theta_K$. Define $\tilde{\epsilon} := \theta_K - U^P_{\phi}(K)$ and observe that $U^P_{\tilde{m}_{\epsilon}} > U^P_{\phi}(K)$ for any $\epsilon < \tilde{\epsilon}$.

A.14 Proof of Remark 2.3

Let us first show that the equilibrium satisfies the Intuitive Criterion with respect to the principal. Suppose $\phi_{\tilde{m}}$ fails the Intuitive Criterion. Then there is a game form m' and a type of the principal, say k, for which the following applies: By hypothesis, in the continuation game that is played after the proposal of m' with point-belief belief about k, k must receive larger payoff than θ_k. Yet, this cannot be the case, since k's largest $(\tilde{\beta}_k = 1)$-feasible payoff is equal to θ_k.

Now consider the agent. Suppose $\phi_{\tilde{m}}$ fails the Intuitive Criterion. Then there exists a type of the agent, say j. By hypothesis, j receives from the non-cooperative play of the FPA with the principal holding degenerate belief concentrated on j strictly larger payoff than in equilibrium. However, we can apply lemma 2.9 to conclude that in this case he would receive payoff $\max_{\check{k} \in \Theta^P} B(\check{k})(j - \theta_{\check{k}})$, which is exactly type j's equilibrium payoff.

Moreover, the allocation induced by $\tilde{m}$ is also ratifiable (see Cramton and Palfrey, 1995). Because agent type J is indifferent between accepting and rejecting $\tilde{m}$, the chosen off-path belief is a credible veto belief, and part (ii) of the definition of ratifiable is satisfied.

A.15 Proof of Proposition 2.5

Suppose collusion is in a SPA. We want to argue that we always can find one equilibrium allocation that is weakly more efficient than the unique equilibrium if collusion is in a FPA.

Recall from the proof of theorem 2.11 that the equilibrium in the FPA can be described by a sequence of weakly increasing threshold-types, $\{j_{\tilde{m}}(k)\}_{k=0}^{K}$, such that principal type k receives the right against any agent type strictly below $j_{\tilde{m}}(k)$.

For the ease of exposition, we focus on the case according to which $\theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k}$ is non-decreasing in k. Let us term this condition (R). By equations (A.16) and (A.17), (R) implies that the set $\check{\Theta}^P$ has no holes and principal types not receiving the right with certainty are not pooled. We remark in footnotes which steps of the proof change if this condition fails.

We want to verify that the following allocation is always an equilibrium in the SPA: Define $\{\tilde{j}(k)\}_{k=0}^{K} := \{\min\{j(k), j_{\tilde{m}}(k)\}\}_{k=0}^{K}$, where $j(k)$ is the RSW threshold, defined in the proof of lemma 2.2. Principal type k receives the right if and only if the agent's type realizes below $\tilde{j}(k)$. Because $\{\tilde{j}(k)\}_{k=0}^{K}$ is defined as the minimum of two weakly increasing sequences, it is itself a weakly increasing sequence and therefore induces a weakly increasing, both for the agent and for the principal, allotment policy. Hence, there exists a linear-separable transfer scheme such that the allotment policy can be implemented in an incentive-compatible manner. We choose this scheme such that every principal's type upward adjacent incentive constraint is satisfied with equality. In addition, the scheme consists of a lump-sum payment to the agent, such that the largest agent type receives exactly his outside-option. Since $j_{\tilde{m}}(k), j(k) \geq \theta_k$ for any k, and by the fact that the non-cooperative play of the SPA leads an efficient allotment of the good, every agent types individual rationality constraint is satisfied if and only if it is satisfied for the largest agent type.

Observe, by construction $\tilde{j}(k) \leq j_{\tilde{m}}(k)$ and thus the above described allocation, termed $\tilde{\phi}$, is less inefficient than the unique equilibrium allocation if collusion is in the FPA.

In order to argue that the above described allocation is an equilibrium allocation, it is therefore left to verify that every type of the principal is left with weakly larger utility than the RSW payoff.

This is done in several steps: To begin with, note that the allocation $\phi_{R\tilde{S}W}$ that is the RSW allocation with entry fee (i.e., the agent is asked to pay a lump-sum transfer to the principal such that type J receives payoff $U_D^A(J)$), if there is a pool at the top, leaves every type of the principal with strictly larger payoff than the RSW allocation.

Given this, we show that $\tilde{\phi}$ dominates $\phi_{R\tilde{S}W}$ from the principal's perspective (and if there no pool at the top, we show that ϕ dominates the RSW allocation). To do so, we proceed iteratively: We start with the lowest principal type, say k', with $j(k') > j_{\tilde{m}}(k')$. Note $k' > 1$, since $j(0) = \theta_0 = j_{\tilde{m}}(0)$. We then verify that every type of the principal prefers the allocation $\tilde{\phi}_{k'}$, which differs from $\phi_{R\tilde{S}W}$ only by the fact that $q^P_{R\tilde{S}W}(j, k') = \mathbb{1}[j < j(k')]$ is replaced by $q^P_{\tilde{\phi}_{k'}}(j, k') = \mathbb{1}[j < j_{\tilde{m}}(k')]$, to the allocation $\phi_{R\tilde{S}W}$.[8] We therefore can without loss of generality replace $\phi_{R\tilde{S}W}$ by $\tilde{\phi}_{k'}$. We then fix the lowest type above k', say k'', with $j(k'') > j_{\tilde{m}}(k'')$ and verify that every type of the principal prefers the allocation $\tilde{\phi}_{k''}$, that is defined in a similar manner as $\tilde{\phi}_{k'}$, to the allocation $\tilde{\phi}_{k'}$. Continuing in this fashion, we finally end up at allocation $\tilde{\phi}$ and are certain that the latter allocation leaves every type of the principal with larger payoff than the RSW payoff.

Given k', define the allotment policy $\{q^A_{\tilde{\phi}_{k'}}(j, k), q^P_{\tilde{\phi}_{k'}}(j, k)\} = \{\mathbb{1}[j \geq j_{\tilde{\phi}_{k'}}(k)], \mathbb{1}[j < j_{\tilde{\phi}_{k'}}(k)]\}$, where the sequence of threshold types is given by: $\{j_{\tilde{\phi}_{k'}}(k)\}_{k=0}^K = \{\min\{j(k), j_{\tilde{m}}(k)\}\}_{k=0}^{k'} \cup \{j(k)\}_{k=k'+1}^K$. By definition, $j_{\phi_0}(k) = j(k)$ for all $k \in \Theta^P$.

The allocation ϕ with $\phi \in \{\phi_0, ..., \phi_K\}$, then reads:

$$(q^A_\phi(j,k), t^A_\phi(j,k)) = \begin{cases} (1, h^A_\phi(j) + \mathbb{E}_k[h^P_\phi(k)] - h^P_\phi(k) - C) & \text{if j} > j_\phi(k), \\ (0, h^A_\phi(j) + \mathbb{E}_k[h^P_\phi(k)] - h^P_\phi(k) - C) & \text{if } j \leq j_\phi(k). \end{cases}$$

$$(q^P_\phi(j,k), t^P_\phi(j,k)) = \begin{cases} (0, h^P_\phi(k) - \mathbb{E}_k[h^P_\phi(k)] - h^A_\phi(j) + C) & \text{if j} > j_\phi(k), \\ (1, h^P_\phi(k) - \mathbb{E}_k[h^P_\phi(k)] - h^A_\phi(j) + C) & \text{if } j \leq j_\phi(k). \end{cases}$$

Where

$$h^A_\phi(j) = q^A_\phi(j)j - \int_0^j q^A_\phi(v)dv,$$

[8] If condition (R) fails, then types of the principal between $k'', k''' \in \check{\Theta}^P$ with $k'' < k'''$ are pooled according to the play of $\tilde{m}$ and receive the right against every agent type lower than $j_{\tilde{m}}(k''')$. Let k' be the largest principal type in the pool with $j(k) > j_{\tilde{m}}(k'')$. Then, for all other types between k' and k''' it must be necessarily the case that $j(k) > j_{\tilde{m}}(k''')$. We construct $\tilde{\phi}_{k'''}$ that differs from $\phi_{R\tilde{S}W}$ only by the fact that those types between k' and k''' receive the right against every agent type lower than $j_{\tilde{m}}(k''')$.

$$C = \int_0^J [q_D^A(j) - q_\phi^A(j)]dj$$

$$h_\phi^P(k) = q_\phi^P(k)\theta_{k-1} - \sum_{v=1}^{k-1}(\theta_v - \theta_{v-1})q_\phi^P(v), \quad \text{if } k > 1$$

$$h_\phi^P(1) = h_\phi^P(0) = 0.$$

For latter references, observe that for any $k > 0$ the utility reads:

$$U_\phi^P(k) = \sum_{v=1}^{k} F(j_\phi(v))(\theta_v - \theta_{v-1}) + \sum_{v=1}^{K} F(j_\phi(v))[\theta_{v-1} - \frac{1-B(v)}{\beta_v}(\theta_v - \theta_{v-1})]\beta_v$$
$$+ \int_0^J q_\phi^A(j)(j + \frac{F(j)}{f(j)})f(j)dj - \int_0^J q_D^A(j)dj$$

Step 1

Fix the lowest type k such that $\tilde{\phi}_0 \neq \tilde{\phi}_k$, i.e., such that $j(k) > j_{\tilde{m}}(k)$. Let us first consider

$$U_{\tilde{\phi}_0}^P(k') - U_{\tilde{\phi}_k}^P(k') =$$
$$\int_{j_{\tilde{m}}(k)}^{j(k)} \{[\theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k}] - (j + \frac{F(j)}{f(j)})\}f(j)\beta_k dj \tag{A.19}$$

We want to argue that this difference is necessarily negative. Because of the regularity assumption on the agent's type distribution, a sufficient condition is given by:

$$\theta_k + (\theta_k - \theta_{k-1})\frac{B(k-1)}{\beta_k} \leq j_{\tilde{m}}(k) + \frac{F(j_{\tilde{m}}(k))}{f(j_{\tilde{m}}(k))}$$

Now, recall from (A.16) that the left-hand side is equal to $j_{\tilde{m}}(k)$, and thus

$$j_{\tilde{m}}(k) \leq j_{\tilde{m}}(k) + \frac{F(j_{\tilde{m}}(k))}{f(j_{\tilde{m}}(k))},$$

which is true.[9] [10]

[9] If condition (R) fails, we jointly decrease the allotment probability of all types in the pool being defined in the previous footnote (8). That is, for every type between k' and k'''. Doing so, one can establish that for type k''' the following holds $U_{\tilde{\phi}_0}^P(k''') - U_{\tilde{\phi}_{k'''}}^P(k) < 0$ (see the next footnote, 10). The resulting allocation, $\tilde{\phi}_{k'''}$, satisfies the principal's upward adjacent incentive constraints. Therefore it follows by the same arguments as in Step 2, that every principal type in the considered pool prefers $\tilde{\phi}_{k'''}$ to $\tilde{\phi}_0$.

[10] Note,

$$U_{\tilde{\phi}_0}^p(k''') - U_{\tilde{\phi}_k}^p(k''') =$$

$$\sum_{v=k'}^{k'''}(F(j(v)) - F(j_{\tilde{m}}(v)))[(\theta_v - \theta_{v-1}) + \theta_{v-1}\beta_v - (1-B(v))(\theta_v - \theta_{v-1})] - \int_0^J [q_{\tilde{\phi}}^a(j) - q_{\tilde{\phi}_0}^a(j)][j + \frac{F(j)}{f(j)}]f(j)dj =$$

$$\sum_{v=k'}^{k'''}\int_{j_{\tilde{m}}(k''')}^{j(v)}[\theta_v + \frac{B(v-1)}{\beta_v}(\theta_v - \theta_{v-1}) - (j + \frac{F(j)}{f(j)})]f(j)\beta_v dj < \sum_{v=k'}^{k'''}\int_{j_{\tilde{m}}(k''')}^{j(v)}[\theta_v + \frac{B(v-1)}{\beta_v}(\theta_v - \theta_{v-1}) - j_{\tilde{m}}(k''')]f(j)\beta_v dj$$

$$< [F(j(k''')) - F(j_{\tilde{m}}(k''')]\sum_{v=k'}^{k''}[\theta_v + \frac{B(v-1)}{\beta_v}(\theta_v - \theta_{v-1}) - j_{\tilde{m}}(k'')]\beta_v],$$

Step 2

Finally, observe that $U^P_{\tilde{\phi}_k}(k-1) = U^P_{\tilde{\phi}_k}(k) - F(\tilde{j}(k))(\theta_k - \theta_{k-1})$ is larger than

$$U^P_{\phi_0}(k-1) = U^P_{\phi_0}(k) - F(j(k))(\theta_k - \theta_{k-1}),$$

because $U^P_{\tilde{\phi}_k}(k-1) - U^P_{\phi_0}(k-1) > (F(j(k)) - F(\tilde{j}(k))(\theta_k - \theta_{k-1}) > 0$.
Moreover, because $q^P_{\phi_k}(k') = q^P_{\phi_{k-1}}(k')$ for all $k' \leq k-1$,

$$U^P_{\tilde{\phi}_k}(k') - U^P_{\phi_0}(k') = U^P_{\tilde{\phi}_k}(k-1) - U^P_{\phi_0}(k-1) > 0 \quad \forall k' < k-1.$$

Similarly, $U^P_{\tilde{\phi}_k}(k+1) = U^P_{\tilde{\phi}_k}(k) + q^P_{\tilde{\phi}_k}(k+1)(\theta_{k+1} - \theta_k)$ is larger than

$$U^P_{\phi_0}(k+1) = U^P_{\phi_0}(k) + q^P_{\phi_0}(k+1)(\theta_{k+1} - \theta_k),$$

because $q^P_{\tilde{\phi}_k}(k') = q^P_{\phi_0}(k')$ for any $k' > k$. As a consequence,

$$U^P_{\tilde{\phi}_k}(k') - U^P_{\phi_0}(k') = U^P_{\tilde{\phi}_k}(k) - U^P_{\phi_0}(k) > 0 \quad \forall k' > k.$$

Since the same arguments as above apply for any k such that $j(k) > j_{\tilde{m}}(k)$, we consider proposition 2.5 as proven.

A.16 Proofs of Section 2.5.1: The Intuitive Criterion

Proof of Lemma 2.13 The next lemma states a sufficient condition such that equilibrium does not fail the CK criterion. That is, if there does not exists any separating menu in the below sense, then the equilibrium does not fail the criterion.

Lemma A.5. *Suppose the equilibrium fails the CK criterion. Then there exist a principal type k and a $(\tilde{\beta}_k = 1)$-feasible menu, say ϕ^S_k, separating principal type k and $k-1$. That is, $U^P_{\phi^S_k}(k-1) \leq U^P(k-1)$, $U^P_{\phi^S_k}(k) > U^P(k)$.*

Proof. For the fixed off-path mechanism m', define a set of side mechanisms $\hat{M}_{m'}$, such that $\hat{m} \in \hat{M}_{m'}$ if and only if any principal type $\underline{k}$ not in $\Xi_1(m')$ receives in the continuation game after the proposal of $\hat{m} \in \hat{M}_{m'}$ payoff lower than equilibrium utility - for any belief $\beta^1 \in B_1(m')$. We know that $\hat{M}_{m'}$ is non empty, since m' is included in this set by the assumption on the failure of the CK criterion.

where the first inequality follows from the regularity assumption on the agent's type distribution and the second from the fact that the principal's virtual valuation is positive.

Recall the definition of k'' in footnote 13 and observe that the last quantity is weakly smaller than zero, because

$$\sum_{v=k'}^{k'''}[\theta_v + \frac{B(v-1)}{\beta_v}(\theta_v - \theta_{v-1}) - j_{\tilde{m}}(k'')]\beta_v \leq \sum_{v=k''+1}^{k'''}[\theta_v + \frac{B(v-1)}{\beta_v}(\theta_v - \theta_{v-1}) - j_{\tilde{m}}(k''')]\beta_v = 0$$

$$\iff B(k''')(j_{\tilde{m}}(k''') - \theta_{k'''}) = B(k'')(j_{\tilde{m}}(k''') - \theta_{k''}) = 0,$$

where the first inequality follows from the fact that for any $k \in \{k''+1, ..., k'''-1\}$ it holds that $j_{\tilde{m}}(k''') < \theta_k + \frac{B(k-1)}{\beta_k}(\theta_k - \theta_{k-1})$ and the last equality is satisfied by the definition of $j_{\tilde{m}}(k''')$.

Define for all $k \in \Theta^P$ the secured payoff $\underline{U}^P_{m'}(k)$, the following way:

$$\underline{U}^P_{m'}(k) := \underset{m \in \hat{M}_{m'}}{sup} \min_{\tilde{\beta} \in B_1(m'), U^P_m(k) \in M(k, \tilde{\beta}, m)} U^P_m(k),$$

Next, consider the sequence of problems, defined for all $\check{k} \in \Theta^P$ as:

$$\max_{\phi} U^P_{\phi}(k) \qquad (P)_{\check{k}}$$

$$(AIC)^j_k, (AIR)^j_k, (PIC)_k \ \ \forall j \in \Theta^A, \forall k \in \Xi_1(m'),$$
$$U^P_{\phi}(k) \le U^P(k) \ \ \forall k \notin \Xi_1(m'),$$
$$(BB)_{j,k},$$

where the constraints refer to the agent's incentive-compatibility and individual rationality constraints - imposed at the ex-post level given any $k \in \Xi_1(m')$ -, and the principal's interim incentive constraint if his type is in the set $\Xi_1(m')$. Otherwise the principal's payoff induced by ϕ is required to be weakly less than his equilibrium utility.

Firstly, note that there exists an allocation satisfying the constraints, namely ϕ^{RSW}: It satisfies $(PIC)_k, (AIC)^j_k, (AIR)^j_k \ \forall (j,k) \in \Theta$ and thus $\forall j \in \Theta^A, \forall k \in \Xi_1(m')$. Moreover, since $U^P(k)$ is an equilibrium payoff $U^P_{RSW}(k) \le U^P(k) \ \ \forall k \in \Theta^P$ and thus $\forall k \notin \Xi_1(m')$.

Secondly, denote the solution to $(P)_k$ by ϕ_k and define $\phi' : \phi'(\cdot, k) = \phi_k(\cdot, k)$. Then observe that, by construction, ϕ' is feasible in any $(P)_k$ and $U^P_{\phi'}(k) = U^P_{\phi_k}(k)$.

Thirdly, note that $U^P_{\phi'}(k) = \underline{U}^P_{m'}(k)$:

Suppose first that $U^P_{\phi'}(k) > \underline{U}^P_{m'}(k)$. This contradicts the definition of $\underline{U}^P_{m'}(k)$. By proposing a convex combination between ϕ^o[11] and ϕ' principal type k can secure himself for any $\tilde{\beta} \in B_1(m')$ a payoff strictly larger than $\underline{U}^P_{m'}(k)$. Moreover, since ϕ^o induces any type of the principal a strictly smaller payoff than the RSW allocation, it follows that the direct revelation mechanism implementing the convex combination between ϕ^o and ϕ' is a feasible side mechanism, i.e., an element of $\hat{M}_{m'}$.

Now suppose that $U^P_{\phi'}(k) < \underline{U}^P_{m'}(k)$. We aim to derive a contradiction. As first step, we define the following set:

$$H_{m'}(k, \tilde{\beta}) := \{U^P_{\phi}(k) | \exists \phi :$$
$$(AIC)^j_{\tilde{\beta}}, (AIR)^j_{\tilde{\beta}}, (PIC)_k \ \ \forall (j,k) \in \Theta^A \times supp(\tilde{\beta}), U^P_{\phi}(k) \le U^P(k) \ \ \forall k \notin \Xi_1(m')\}.$$

In words, the set $H_{m'}(k, \tilde{\beta})$ fixes a belief and varies over all $m \in \hat{M}_{m'}$. Note first, the revelation principle implies that the continuation game starting after an offer of any $m \in \hat{M}_{m'}$ together with $\tilde{\beta} \in B_1(m')$ induces an allocation ϕ such that $U^P_{\phi}(k)$ is an element of $H_{m'}(k, \tilde{\beta})$. We abuse notation a bid and call any such allocation $\tilde{\beta}$-feasible.

Also note that the quasilinear environment implies that $H_{m'}(k, \tilde{\beta})$ is a convex set.

[11]Recall that ϕ^o is defined as allocation that satisfies ex-post incentive compatibility and individual rationality with strict inequality.

Moreover, from the definition of the secured payoff it follows that for any $\tilde{\beta} \in B_1(m')$ it is the case that $\underline{U}^P_{m'}(k) \in H_{m'}(k, \tilde{\beta})$: $\underline{U}^P_{m'}(k) > U^P_\phi(k)$ for any $\tilde{\beta}'$-feasible allocation ϕ, some $\tilde{\beta}' \in B_1(m')$ and some k implies a contradiction to the fact that $\underline{U}^P_{m'}(k)$ is defined as the value of a minimization problem with respect to the belief. $\underline{U}^P_{m'}(k) < U^P_\phi(k)$ for any $\tilde{\beta}'$-feasible allocation ϕ, some $\tilde{\beta}' \in B_1(m')$ and some k cannot be the case since by hypothesis $U^P_{\phi'}(k) < \underline{U}^P_{m'}(k)$ and $U^P_{\phi'}(k) \in H_{m'}(k, \tilde{\beta})\ \ \forall \tilde{\beta} \in B_1(m')$. Since $H_{m'}(k, \tilde{\beta})$ is a convex set, the claim follows.

We thus note that there exists a payoff, being feasible for any $\tilde{\beta} \in B_1(m')$. Invoking the revelation principle, we observe:

$$\underline{U}^P_{m'}(k) \in \{U^P_\phi(k) | \exists \phi :$$

$$(AIC)^j_k, (AIR)^j_k, (PIC)_k\ \ \forall (j,k) \in \Theta^A \times \Xi_1(m'), U^P_\phi(k) \le U^P(k)\ \ \forall k \notin \Xi_1(m')\}.$$

The hypothesis $\underline{U}_{m'}(k) > U^P_{\phi'}(k)$ thus implies a contradiction. There is a game form m' that is feasible in $(P)_k$ and leads a higher payoff than the solution value of program $(P)_k$. Finally, we want to show that the allocation ϕ' implies the existence of a separating menu.

First, take any principal type in $k' \in \Xi_2(m')$ and note that

$$U^P_{\phi'}(k') = \underline{U}_{m'}(k') \ge \min_{\tilde{\beta} \in B_1(m')} M(k', \tilde{\beta}, m') > U^P(k')$$

The first inequality follows from the fact that m' is not necessarily the game form that induces the supremum,[12] and the last inequality follows from the failure of the CK criterion.

Now fix the lowest type $k' \in \Xi_2(m')$ and define the set

$$\Xi_3(k', m') := \{k \in \{0, ..., k'-1\} | k \in \Xi_1(m'), k+1 \in \Xi_3(k', m')\} \cup k'$$

In words, $\Xi_3(k', m')$ is a subset of types in $\Xi_1(m')$ with no "holes". That is, k' is the largest element of $\Xi_3(k', m')$. $k'-1 \in \Xi_3(k', m')$ if and only if $k'-1 \in \Xi_1(m')$ and $k'-2 \in \Xi_3(k', m')$ if and only if $k'-2, k'-1 \in \Xi_1(m')$ and so on ... .

If $k'-1$ [13] is not in $\Xi_1(m')$, we have found a separation menu by the construction of the maximization problem $(P)_{k'}$. Hence, assume that $k'-1$ is in $\Xi_1(m')$.

If for any type $k \in \Xi_3(k', m')$ it is the case that $U^P(k) \ge U^P_{\phi'}(k)$ we are done, as we can simply take the highest type, say k'', with this feature. Then we know that there exists a separation menu for $k''+1$.

We thus fix the highest principal type not in $\Xi_3(k', m')$. This type is strictly below $k'-1$. Denote him by k'''.[14] By definition of $\Xi_3(k', m')$, k''' cannot be in $\Xi_1(m')$.

[12] Recall that $M(k', \tilde{\beta}, m')$ is the continuation payoff of principal type k' in the continuation game, following m' under belief $\tilde{\beta} \in B_1(m')$.

[13] Such a type exists, since we can assume without loss of generality that $k = 0$ is not in $\Xi_2(m')$: $k = 0$ cannot be in $\Xi_2(m')$ since in this instance, type 0 would need to profit, in particular, if the agent held degenerate belief concentrated on type 0. By the features of the RSW allocation, the lowest principal type cannot receive higher utility than his RSW payoff, when the agent has a degenerate belief concentrated on this type.

[14] Such a type exists, since we can assume without loss of generality that $k = 0$ is not in $\Xi_2(m')$, as argued in footnote 13 above. Moreover, if $k = 0$ is in the set $\Xi_3(k', m')$ (and thus in the set $\Xi_1(m')$), it cannot be the case that all types in $\Xi_1(m')$ prefer ϕ' to the equilibrium mechanism: Otherwise all types

By construction, we therefore have found a separating menu between the lowest principal type in $\Xi_3(k', m')$, $k''' + 1$, and k'''. □

We now turn to the necessary part. Suppose that some principal type k has a separation menu. We aim to construct a deviation mechanism that makes the equilibrium allocation failing the CK criterion.
In order to proceed, consider first the following sequence of linked menu-maximization problems, $(P4)_k$.

$$\max_{\phi(\cdot,0)} U^P_{\phi(\cdot,0)}(0) \qquad (P4)_0$$

such that $\forall j \in \Theta^A$:

$$(AIR)^j_0, (AIC)^j_0$$

$$(BB)_{j,0} \qquad t^A_{\phi(\cdot,0)}(j) + t^P_{\phi(\cdot,0)}(j) \geq 0,$$

and for any $k = 1, ..., K$ and given $\phi(\cdot, 0), ..., \phi(\cdot, k-1)$,

$$\max_{\phi(\cdot,k)} U^P_{\phi(\cdot,k)}(k) \qquad (P4)_k$$

such that $\forall j \in \Theta^A$:

$$(AIR)^j_k, (AIC)^j_k$$

$$(PIC)^+_{k-1} \qquad \{\max U^P_{\phi(\cdot,k-1)}(k-1), U^P(k-1)\} \geq U^P_{\phi(\cdot,k)}(k-1)$$

$$(BB)_{j,k} \qquad t^A_{\phi(\cdot,k)}(j) + t^P_{\phi(\cdot,k)}(j) \geq 0,$$

where each $(P4)_k$ defines an optimal - from principal type k's point of view - menu offer to the agent, such that $k-1$'s upward adjacent incentive constraint ($(PIC)^+_{k-1}$) is satisfied. Each $(P4)_k$ is subject to agent's ex-post individual rationality and incentive-compatibility constraints, given k.
Note the similarity between $(P4)_k$ and the sequence of linked-(menu)-maximization problems $(\tilde{P1})_k$, which we solved to characterize the RSW allocation. In both programs, the agent's individual rationality and interim incentive-compatibility constraints are imposed for every type of the principal at the ex-post level. According to $(\tilde{P1})_k$ the principal's upward adjacent incentive constraints bind. The problems in program $(P4)_k$ are still linked but subject to a weakened (PIC).
In order to construct the deviation mechanism, we solve the class of maximization problems $(P4)_k$ by starting with the lowest principal type. We set k' equal to the first k such that $U^P(k) \leq U^P_{\phi(\cdot,k)}$ and replace $(PIC)^+_{k'}$ by the constraint $U^P_{\phi(\cdot,k')}(k') \geq U^P_{\phi(\cdot,k'+1)}(k')$. Clearly, the assumption of the existence of a separation menu implies that k' is well-defined. Finally note, if for any k it holds that $U^P_{\phi(\cdot,k)}(k) \geq U^P(k)$ then $U^P_{\phi(\cdot,k+1)}(k+1) \geq U^P_{\phi(\cdot,k)}(k+1)$ as the agents outside option is eased in the type of the principal.
Given the similarity between $(P4)_k$ and $(\tilde{P1})_k$ it should be clear (we omit the proof) that we can describe $\phi(\cdot, k)$ by a binary 0,1 allotment policy in which agent types

from $k = 0$ to k' would weakly (and at least k' strictly) prefer some ϕ satisfying $(AIR)^j_k$,$(AIC)^j_k$ and $(PIC)_k$ for all $k \in \{0, ..., k'\}$: a contradiction to the fact that the equilibrium payoff weakly dominates the RSW payoff. I.e., ϕ' would be the RSW allocation. But by assumption, k' prefers the utility induced by ϕ' strictly to his equilibrium payoff. This leads a contradiction, since $U^P(k') \geq U^P_{RSW}(k')$.

below some threshold, $\tilde{j}(k)$ with $\tilde{j}(k) \geq \theta_k$, receive the compensation $\tilde{j}(k) - \theta_k$. Types above receive the right at price $\theta_{k'}$. Similar to the RSW allocation, a lump-sum transfer might be paid to the agent if the principal cannot use the allotment probability to separate from his downward-adjacent neighbor. That is, there might exists a pool at the top of principal types with lowest element indicated by $\underline{K}$.

We now construct a side mechanism, the deviation mechanism, that makes the equilibrium failing the CK criterion. In the deviation mechanism the agents message space is augmented. Together with his type the agent reports a number $d \in \{0,1\}$. Moreover, define the function $l : \{0,1\} \times \Theta^P \to \mathbb{R}^2$ that elicits the agent's belief. If the agent reports $d = 0$ he receives a payment, D^A, if and only if the principal reports to have type below k'

$$(l^A(0,\hat{k}), l^P(0,\hat{k})) = \begin{cases} (-D^A, D^P) & \text{if } \hat{k} < k' \,, \\ (\epsilon, D^P) & \text{if } \hat{k} \geq k' \,. \end{cases}$$

$$(l^A(1,\hat{k}), l^P(1,\hat{k})) = \left\{ \begin{array}{ll} (0,0) & \forall \hat{k} \in \Theta^P \,. \end{array} \right.$$

D^A is chosen such that no agent type rejects the deviation mechanism. That is $-D^A > t^P(0) + U^A_D(J)$. D^P is chosen such to assure a non-negative budget surplus, i.e., $D^P = -D^A$. Recall, the objects $t^P(\hat{k}), q^P(\hat{j},\hat{k})$ refer to the considered equilibrium allocation. Given this, the deviation mechanism is defined by the outcome function $\phi^0_{CK}(\hat{j},\hat{k},d)$:

$\hat{k} < k'$:

$$(q^A_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^A_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (0, \max\{t^P(\hat{k}),0\} - l^A(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

$$(q^P_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^P_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (q^P(\hat{j},\hat{k}), t^P(\hat{k}) + l^P(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

$\hat{k} \geq k'$:

$$(q^A_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^A_{\phi^0_{CK}}(\hat{j},\hat{k},d) = \begin{cases} (0, \theta_{\hat{k}} - \tilde{j}(\hat{k}) + l^A(d,\hat{k})) & \text{if } \hat{j} \leq \tilde{j}(\hat{k}) \,, \\ (1, \theta_{\hat{k}} + l^A(d,\hat{k})) & \text{if } \hat{j} > \tilde{j}(\hat{k}) \,. \end{cases}$$

$$(q^P_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^A_{\phi^0_{CK}}(\hat{j},\hat{k}) = \begin{cases} (1, \theta_{\hat{k}} - \tilde{j}(\hat{k}) + l^P(d,\hat{k})) & \text{if } \hat{j} \leq \tilde{j}(\hat{k}) \,, \\ (0, -\theta_{\hat{k}} + l^P(d,\hat{k})) & \text{if } \hat{j} > \tilde{j}(\hat{k}) \,. \end{cases}$$

$\hat{k} = \underline{K}$

$$(q^A_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^A_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (0, -\theta_{\underline{K}} + l^A(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

$$(q^P_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^P_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (1, -(\theta_{\underline{K}-1} - U^P_{\phi^0_{CK}}(\underline{K}-1)) + l^P(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

$\hat{k} > \underline{K}$

$$(q^A_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^A_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (0, l^A(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

$$(q^P_{\phi^0_{CK}}(\hat{j},\hat{k},d), t^P_{\phi^0_{CK}}(\hat{j},\hat{k},d)) = \left\{ \begin{array}{ll} (1, -(\theta_{\underline{K}-1} - U^P_{\phi^0_{CK}}(\underline{K}-1)) + l^P(d,\hat{k})) & \forall \hat{j} \,. \end{array} \right.$$

Finally, replace ϕ^0_{CK} by the convex combination between ϕ^o, being defined in lemma 2.3 as allocation that satisfies the agent's incentive constraints and individual rationality constraints strictly, and ϕ^0_{CK} with mass approaching unity on the latter outcome function.

Denote the resulting allocation by ϕ_{CK} and the mechanism implementing it, upon unanimous ratification of all agent types, by m_{CK}.

Lemma A.6. *Whenever there exists a deviation mechanism, the equilibrium fails the CK criterion.*

Proof. By construction, m_{CK} is accepted by any agent type for any belief about the principal. Moreover, for any belief about the principal m_{CK} has a unique and truthful, concerning the bidders' type reports, equilibrium. Moreover, the set of types that profit from m_{CK} in comparison to equilibrium for some belief, $\Xi_1(m_{CK})$, includes at least k' and no type below k'. Finally, for any belief with support only on types in the set $\Xi_1(m_{CK})$, every type of the agent reports $d = 1$ and k' receives larger payoff than in equilibrium.

□

Proof of Proposition 2.6 We start by verifying the conditions for the (non-)existence of the separation menu, defined in the text. Having established this result, the proposition 2.6 is a direct consequence.
We begin with the necessary conditions:

Lemma A.7. *If k pays a strictly positive subsidy and $k-1$'s upward adjacent incentive constraint is not satisfied with equality, then there exists a separating menu.*

Proof. The following menu is a separation menu. Take $q^P_{\phi(\cdot,k)} = q^P(k)$, $t^P_{\phi(\cdot,k)} = t^P(k) - \epsilon$. Where $\epsilon > 0$ is chosen such that $k-1$ does not prefer the allocation induced by the menu to his equilibrium allocation and the agent's ex-post individual rationality constraints, given type k, are satisfied. Such an ϵ exists, since when setting it equal to zero, both $k-1$'s incentive and the agent's individual rationality constraints are satisfied with strict inequality. □

Lemma A.8. *If for any principal type k with $q^P(k) < 1$ there does not exist a separating menu, then k receives a subsidy.*

Proof. Suppose to the contrary of what we want to show that k pays a strictly positive subsidy. We want to verify that in this instance type k has a separation menu.
By lemma A.7 assume without loss of generality that $k-1$'s upward adjacent incentive constraint is satisfied with equality. For some $\epsilon, \delta > 0$, construct the following menu: $q^P_{\phi(\cdot,k)} = q^P(k) + \epsilon$ and $t^P_{\phi(\cdot,k)} = t^P(k) + \theta_{k-1}\epsilon + \delta$.
We choose ϵ positive, but arbitrarily small, such that this menu satisfied $(AIR)_k$. This is feasible, since $q^P(k) + \epsilon < 1$ and $(AIR)_k$ is satisfied with strict inequality for $\epsilon = 0$ (by the assumption of the subsidy). This feature, or feasibility, is preserved for ϵ arbitrarily small and also for any positive δ. For $0 < \delta < (\theta_k - \theta_{k-1})\epsilon$ type $k-1$ does not profit from this menu, but k does so.
Hence, we have found a separating menu.

□

Next we prove the sufficient conditions:

Lemma A.9. *Whenever (i) $k-1$'s upward adjacent incentive constraint is satisfied with equality, (ii) k receives a subsidy and (iii) $q^P(k) \geq F(\theta_k)$, then there does not exist a separating menu.*

Proof. Suppose to the contrary that there exists a separating menu. Since $k-1$'s upward adjacent incentive constraint is satisfied with equality in equilibrium, there only can exists a separating menu if $q^P_{\phi^S_k} > q^P(k)$: Suppose to the contrary $q^P_{\phi^S_k} \leq q^P(k)$. In this instance $U^P_{\phi^S_k}(k-1) = U^P_{\phi^S_k}(k) - q^P_{\phi^S_k}(\theta_k - \theta_{k-1}) \geq U^P(k) - q^P(k)(\theta_k - \theta_{k-1}) = U^P(k-1)$, which is a contradiction to the definition of the separating menu. However, $q^P_{\phi^S_k} > q^P(k)$ together with the hypothesis that k does not receive a subsidy implies $U^P_{\phi^S_k}(k) < U^P(k)$: First, observe that the hypothesis that k is subsidized implies that the equilibrium menu, ϕ^E_k, induces k a payoff $U^P_{\phi^E_k}(k)$, strictly smaller than $U^P(k)$. However, as $q^P_{\phi^S_k} > q^P(k) \geq F(\theta_k)$ the allocation is already (weakly) inefficient, given type k. As a consequence, when increasing the inefficiencies, that is, increasing $q^P_{\phi^S_k}$, the payoff principal type k receives, $U^P_{\phi^S_k}(k)$, furthermore decreases. □

Lemma A.10. *Fix a type k such that $q^P(k) = 1$. Suppose $k-1$'s upward adjacent constraints is satisfied with equality, then there does not exist a separating menu.*

Proof. Along the same lines as in the proof of lemma A.9 it can be established that if $k-1$'s upward adjacent constraint is satisfied, a separation menu must feature $1 = q^P(k) < q^P_{\phi^S_k}$. Clearly, a contradiction. □

With these above observations we can proof the assertions of proposition 2.6.

Lemma A.11. *An equilibrium survives the CK criterion only if the induced allotment policy is inefficient in favor of the principal.*

Proof. Assume first that $K \geq J$. Suppose that $q^P(k) = F(\theta_k)$ for any k, and assume that the equilibrium allocation satisfies the CK criterion. We now argue that there is at least one type k with $q^P(k) < 1$ paying a strictly positive subsidy. First, define k' as the highest type with $q^P(k') < 1$. We know that the upward adjacent incentive constraint of k' is satisfied with equality, because $k'+1$ pays a strictly positive subsidy. To see this, note that the allocation is efficient and $q^P(k'+1) = 1$. Therefore $k'+1$'s equilibrium menu specifies the transfer $t^P_{\phi^E_{k'+1}} = 0$. Furthermore, it must be the case that $t^P(k'+1) > 0$. Otherwise it would follow that $U^P(k') \geq \theta_{k'}$ for all types of the principal and with strict inequality for some. This is the case, because $\theta_k - U^P(k)$ is increasing in k, as $(\theta_k - \theta_{k-1}) - (U^P(k+1) - U^P(k)) \geq (\theta_k - \theta_{k-1}) - (\theta_k - \theta_{k-1})q^P(k)) > 0$. But this implies that agent's individual rationality is violated, by accounting logic: It follows by (AIR) that $\sum_{k=0}^K U^P(k) \leq \sum_{k=0}^K \theta_k$, since the right-hand side is the upper bound of the left-hand side given any β-feasible allocation.[15] Hence, $k'+1$ pays a subsidy and thus, by lemma A.7, the upward adjacent incentive constraint of k' is satisfied with equality.

[15]Suppose we want to maximize $\sum_{k=0}^K U^P(k)$, subject to $(AIR)_k, (AIC)_k \;\; \forall k \in \Theta^P$, but drop the (PIC) constraint. In this instance every element of the sum takes the value θ_k: Each principal can offer a menu that sells the right to the agent at price θ_k and otherwise allots the right to the principal at zero compensation. This mechanism implies utility of θ_k for each principal type k. Since the implied allotment policy is efficient and every agent type receives exactly his outside option, it follows that this is also the solution when imposing the weaker (AIR) and (AIC) constraints. Therefore we constructed an upper bound on the sum of principal (equilibrium) utilities.

We conclude that the following is true

$$\theta_{k'} - t^P(k'+1) = U^P(k')$$

Hence, $U^P(k') < \theta_{k'}$. But this implies that k' pays a strictly positive subsidy, since his equilibrium menu induces payoff $\theta_{k'}$.
By lemma A.8 it thus follows that there exists a separating menu for type k' and we thus have derived a contradiction to the hypothesis that the allocation satisfies the CK criterion.
Now suppose that $K < J$, i.e., $q^P(K) < 1$. Clearly, K pays a strictly positive subsidy: Otherwise he would receive utility $U^P(K) = \theta_K$, which implies, by a similar argument as above, a contradiction to (AIR). Therefore there exists a separation menu for K. □

Lemma A.12. *The RSW allocation survives the CK criterion. Moreover, whenever* $q^P_{RSW}(K) = 1$, *the allocation termed "RSW with entry fee" survives the CK criterion.*

Proof. The RSW allocation survives the CK criterion:
If the RSW allocation is the equilibrium allocation, then, by the characterization of the RSW allocation, the principal's upward adjacent incentive constraints are satisfied with equality. Moreover, all types with $q^P_{RSW}(k) < 1$ pay and receive zero subsidy. Hence, by lemmas A.9 and A.10 there does not exist a separating menu for any principal type.
The RSW allocation with entry fee survives the CK criterion:
If the RSW with entry fee allocation is the equilibrium allocation, then, by the characterization of the RSW allocation, the principal's upward adjacent incentive constraints are satisfied with equality. Moreover, by the characterization of the RSW allocation, it follows that all types with $q^P_{RSW}(k) < 1$ receive a subsidy: The transfer according to the RSW with entry fee allocation is strictly smaller than according to the RSW allocation, i.e., to the equilibrium menu. Thus, there does not exist a separation menu for any type of the principal. □

Lemma A.13. *Whenever* $q^P_{RSW}(K) < 1$ *only the RSW allocation survives the CK criterion*

Proof. Suppose there exists some equilibrium allocation satisfying the CK criterion. In the following we show that it is necessarily the RSW allocation.
First note, all types of the principal receive a weakly positive subsidy. Fix any type k, such that $q^P(k) = 1$. Since $q^P_{RSW}(k) < 1$ and $U^P(k) \geq U^P_{RSW}(k)$, the fact that k receives zero subsidy according to the RSW allocation immediately implies that k receives a strictly positive subsidy in equilibrium. A type k with $q^P(k) < 1$ cannot pay a strictly positive subsidy, since otherwise lemma A.8 implies the existence of a separation menu.
This immediately implies that every type receives zero subsidy. If at least one type receives a strictly positive subsidy, the agent's interim individual rationality constraint would be violated by account logic. But if every principal type receives zero subsidy, then $q^P(k) \leq q^P_{RSW}(k)$, as otherwise $U^P(k) < U^P_{RSW}(k)$.
Moreover, we know that $q^P(k) \geq q^P_{RSW}(k)$. Suppose to the contrary that $q^P(k) < q^P_{RSW}(k)$. We claim that there exists at least one type of the principal being below k that receives a strictly positive subsidy.

We verify this claim by induction and start with $k-1$. If type $k-1$ does not receive a strictly positive subsidy, then incentive compatibility implies that he must receive a strictly larger equilibrium payoff than $U^P_{RSW}(k-1)$. This is the case because the hypothesis $q^P(k) < q^P_{RSW}(k)$ together with the zero subsidy condition imply that $U^P(k, k-1) > U^P_{RSW}(k, k-1) = U^P_{RSW}(k-1, k-1)$. Hence, $q^P(k-1) < q^P_{RSW}(k-1)$. By a similar argument one can establish that $U^P(k-2) > U^P_{RSW}(k-2)$ and $q^P(k-2) < q^P_{RSW}(k-2)$. Continuing this reasoning for all $k > 0$, we eventually we end up by at $k = 0$. By incentive compatibility, we thus know that $U^P(0) > U^P_{RSW}(0)$. However, according to the RSW allocation, principal type 0 receives the largest feasible payoff, given the agent knew the type of the principal. Consequently, this implies that type 0 receives a strictly positive subsidy in equilibrium. We thus conclude that $q^P(k) = q^P_{RSW}(k)$ for all k. □

A.17 Proof of Proposition 2.7

First observe that the only equilibrium that satisfies the CK criterion is the RSW allocation.
The proof consists of several steps. We first assume that it is without loss of generality to assume that the $r \geq \theta_0$ and then verify this claim in step 4.
We begin by stating the welfare objective:
Step 1:

$$W(\phi^{RSW}, r) = \sum_{k=0}^{K} \mathbb{1}[\theta_k \geq r] F(j_r(k))\theta_k\beta_k + \sum_{k=0}^{K} \mathbb{1}[\theta_k \geq r] \int_{j_r(k)}^{J} jf(j)dj\beta(k) + \sum_{k=0}^{K} \mathbb{1}[\theta_k < r]\beta_k \int_{r}^{J} f(j)jdj$$

At $r \in (\theta_k, \theta_{k+1})$ the derivative of $W(\phi^{RSW}, r)$ with respect to r reads:

$$\frac{dW(\phi^{RSW}, r)}{dr}|_{r\in(\theta_k,\theta_{k+1})} = \sum_{k=0}^{K} \mathbb{1}[\theta_k \geq r]\{\frac{dF(j_r(k))}{dr}\theta_k - \frac{dj_r(k)}{dr} j_r(k) f(j_r(k))\}\beta_k - B(r)f(r)r.$$

At $r = \theta_k$ the welfare change induced by a marginal increase of r reads:

$$W(\phi^{RSW}, r > \theta_k) - W(\phi^{RSW}, r = \theta_k) = dr\frac{dW(\phi^{RSW}, r)}{dr}|_{r>\theta_k, r\to\theta_k} - \beta_k F(r)r.$$

Step 2:

We first consider $\frac{dW(\phi^{RSW},r)}{dr}|_{r>\theta_0,r\to\theta_0} + \beta_0 r f(r)$. Observe that for any k>0: $\frac{dF(j_r(k))}{dr}\tilde{\theta}_k - j_r(k)f(j_r(k))\frac{dj_r(k)}{dr} \geq 0$, whenever $\frac{dj_r(k)}{dr} \leq 0$, since $\theta_k \leq j_r(k)$. We assume for the moment that this is the case.
Making use of $\frac{dF(j_r(k))}{dr} = f(j_r(k))\frac{dj_r(k)}{dr}$, we thus can bound $\frac{dW(\phi^{RSW},r)}{dr}|_{r>\theta_0,r\to\theta_0} + \beta_0 r f(r)$ from below by:

$$(\theta_1 - j_r(1))\beta_1 f(j_r(1))\frac{dj_r(1)}{dr}$$

This quantity is strictly positive whenever $\frac{dj_r(1)}{dr} < 0$.

In order to see why this is the case, consider the RSW allocation. $j_r(1)$ is determined by the equation:

$$-F(j_r(1))(j_r(1)-\theta_1)+(1-F(j_r(1))(\theta_1-r)=0$$

$$\iff -F(j_r(1))(j_r(1)-r)+\theta_1-r=0.$$

By the implicit function theorem this defines:

$$\frac{dj_r(1)}{dr}=\frac{(1-F(j_r(1))}{-(F(j_r(1))+f(j_r(1))(j_r(1))-r)}<0,$$

which verifies the claim.
Now fix any $k>0$ and note that $j_r(k)$ is determined by:

$$-F(j_r(k))(j_r(k)-\theta_k)+(1-F(j_r(k))(\theta_k-r)+F(j_r(k))(\theta_{k-1}-r)=U^P_{RSW}(k-1)$$

$$\iff \theta_k-r-F(j_r(k))(j_r(k)-\theta_{k-1})=\theta_{k-1}-r-F(j_r(k-1))(j_r(k-1)-\theta_{k-1})$$

$$\iff \theta_k-\theta_{k-1}-F(j_r(k))(j_r(k)-\theta_{k-1})+F(j_r(k-1))(j_r(k-1)-\theta_{k-1})=0.$$

Now, by the implicit function theorem

$$\frac{dj_r(k)}{dj_r(k-1)}=\frac{f(j_r(k-1))(j_r(k-1)-\theta_{k-1})+F(j_r(k-1))}{f(j_r(k))(j_r(k)-\theta_k)+F(j_r(k))}>0.$$

By induction it does follows, that $\frac{dj_r(1)}{dr}<0$ and thus $\frac{dj_r(k)}{dr}<0$ for all $k\in\Theta^P$.
Step 3:
Now we want to argue that a non-trivial reserve price dominates a trivial reserve price in terms of welfare. Therefore, compare the difference in welfare:

$$lim_{r\to\theta_0,r>\theta_0}W(\phi^{RSW},r)-W(\phi^{RSW},r=\theta_0)=dr\frac{dW(\phi^{RSW},r)}{dr}|_{r\to\theta_0,r>\theta_0}-\beta_0F(\theta_0)\theta_0$$

$$=dr[\sum_{k=1}^{K}\{\frac{dF(j_r(k))}{dr}\theta_k-\frac{dj_r(k)}{dr}j_r(k)f(j_r(k))\}\beta_k-B(0)f(\theta_0)\theta_0]-\beta_0F(\theta_0)\theta_0.$$

As argued above $\sum_{k=1}^{K}\{\frac{dF(j_r(k))}{dr}\theta_k-\frac{dj_r(k)}{dr}j_r(k)f(j_r(k))\}\beta_k>0$. Hence, whenever $F(\theta_0)=0$ and $B(0)=\beta_0\to0$, then a reserve price $r>\theta_0$ welfare dominates a reserve price $r=\theta_0$.
Step 4:
Suppose the lowest type of the principal (and of the agent) is larger than 0, that is, $\theta_0>0$. Then a reserve price of $r=\theta_0$ leads the same welfare as a reserve price of $r<\theta_0$.
Firstly, observe that a reserve price of $r=\theta_0$ does not exclude a strictly positive mass of bidder types. Hence, whenever $r=\theta_0$ implies the same RSW thresholds as $r<\theta_0$, both reserve prices imply the same welfare. To see that this is the case, suppose first that $r<\theta_0$. In this instance $U^P_{RSW}(0)=\theta_0-r$ and, given this, $j_r(1)$ is defined as solution to:

$$F(j_r(1))(\theta_0-r)-F(j_r(1))(j_r(1)-\theta_1)+(1-F(j_r(1)))(\theta_1-r)=\theta_0-r$$

$$\iff \theta_1 - \frac{F(j_r(1))}{1 - F(j_r(1))}(j_r(1) - \theta_1) = \theta_0.$$

In contrast, if $r = \theta_0$ then $U^P_{RSW}(0) = 0$ and with this $j_r(1)$ solves:

$$F(j_r(1))0 - F(j_r(1))(j_r(1) - \theta_1) + (1 - F(j_r(1)))(\theta_1 - r) = 0$$

$$\iff \theta_1 - \frac{F(j_r(1))}{1 - F(j_r(1))}(j_r(1) - \theta_1) = r = \theta_0.$$

As a consequence $j_{r=\theta_0}(1) = j_{r<\theta_0}(1)$. By similar arguments as in step 2 above it follows that $j_{r=\theta_0}(k) = j_{r<\theta_0}(k) \;\; \forall k \in \Theta^P$ which was to be shown.

Appendix B

Appendix Chapter 3

B.1 Details on Rewriting the Problem

Full Participation. Full participation is a consequence of the fact that litigation utility is convex in beliefs and Proposition 2 from Celik and Peters (2011).

Lemma B.1. *It is without loss of generality to assume full participation in the optimal mechanism.*

Value of Vetoing. Any off-path belief structure that satisfies the intuitive criterion leads to a player independent value $V_i(k)$, which is

$$V(l) = (1-p)\frac{\kappa - 1}{\kappa}, \text{ and } V(h) = 0.$$

Given the constant outside option, the channels identified by Celik and Peters (2011) and Cramton and Palfrey (1995) are not present in our model as off path beliefs are less important.

Reduced Form Problem à la Border (2007). We reduce the problem by replacing the settlement shares, X_i, by the settlement values, z_i. For any given matrix of breakdown probabilities, G, this reduction is possible if and only if each settlement share is both individually feasible (condition (F), below) and ex-post implementable (condition (EPI), below). The following lemma states these conditions. With some abuse of notation, let $p(m)$ be the ex-ante probability that player i is of type m, that is $p(l) = p$ and $p(h) = 1 - p$.

Lemma B.2. *For every message $m \in \{l, h\}$, let $m^c := \{k \in \{l, h\} | k \neq m\}$, and fix some feasible G and $z_i \geq 0$ for every i. Then there exists an ex-post feasible X_i that implements z_i if and only if the following constraints are satisfied:*

- $\forall \{m, n\} \in \{h, l\}^2$:
$$p(m)z_i(m) + p(n)z_{-i}(n) \leq \qquad (EPI)$$
$$1 - Pr(\Gamma) - (1 - \gamma(m^c, n^c))p(m^c)p(n^c)$$

- $\forall m \in \{h, l\}$ *and* $i = 1, 2$:
$$z_i(m) \leq 1 - \gamma_i(m) \qquad (IF)$$

Moreover, if $\gamma_i(l) \geq \gamma_i(h)$ then $z_i(l) \leq 1 - \gamma_i(l)$ and (IC_i^l) imply equation (IF).

Note that a necessary condition for individual feasibility (IF) is that it holds in expectations, that is the weighted sum of settlement values cannot exceed the probability of successful ADR,

$$\sum_{i\in\{1,2\}} \sum_{m\in\{l,h\}} p(m)z_i(m) \leq 1 - Pr(\Gamma). \tag{AF}$$

The High-Cost's IC and the Low-Cost's PC bind. Next, we eliminate all settlement values with help of the following lemma stating that in the optimal mechanism the high-cost type's incentive constraint and the low-cost type's participation constraint bind for both players.

Lemma B.3. *It is without loss of generality to assume that* (IC_i^h) *and* (PC_i^l) *hold with equality in the optimal mechanism.*

The result is a direct consequence of the different costs. High-cost types care more about settlement than about breakdown. Thus, incentive compatibility requires a large value of settlement, $z_i(h)$, for them. However, there is no reason for the mediator to set $z_i(h)$ too high, as the h-type would never veto ADR. We can express (IC_i^h) as

$$z_i(h) + \gamma_i(h)U_i(h|h) \geq z_i(l) + \gamma_i(l)U_i(h|l). \tag{IC_i^h}$$

If this inequality is strict, the mediator can reduce the value of settlement, $z_i(h)$, without affecting the breakdown probability $Pr(\Gamma)$ or any of the other constraints. Similarly the mediator can reduce the value of settlement, $z_i(l)$, if l-types' participation constraint is not binding, as any negative effect on l-types incentive constraint (IC_i^l) is of second order compared to the positive effect on h-types incentive constraint, (IC_i^h). By readjusting the settlement value for h-types, $z_i(h)$, incentive compatibility for both types can always be guaranteed. The l-types participation constraint is

$$z_i(l) + \gamma_i(l)U_i(l|l) \geq V(l). \tag{PC_i^l}$$

Using (PC_i^l), (IC_i^h) and Lemma B.3 we can eliminate all settlement values, z_i, and express the result only in terms of breakdown valuations, $\gamma_i(m)U_i(k|m)$.

Breakdown Probabilities and Beliefs. Breakdown beliefs $p_i(l|k)$ are a result of breakdown probabilities. The belief that player 1 is type l, given 2 reported m is

$$p_1(l|m) = \frac{p\gamma(l,m)}{p\gamma(l,m) + (1-p)\gamma(h,m)}.$$

Observation B.1. Any $p_i(l|m)$ is homogeneous of degree 0 in G.

Thus, any set of beliefs $p_i(k|m)$ induced by some G is induced by $G' = \alpha G$, too.

Lemma B.4. *Fix any feasible* G *with* $1 \geq \gamma(l,h), \gamma(h,l), \gamma(l,l) \geq 0$ *and define*

$$q_i(m) := \frac{p}{1-p}\frac{1-p_i(l|m)}{p_i(l|m)}.$$

Then the induced information structure $P > 0$ satisfies:

$$\gamma(h,l) = q_1(l)\gamma(l,l) \leq 1 \qquad \gamma(l,h) = q_2(l)\gamma(l,l) \leq 1; \tag{C}$$
$$\gamma(h,h) = q_2(h)q_1(l)\gamma(l,l) \leq 1 \qquad q_2(h)q_1(l) = q_1(h)q_2(l),$$

where the last equation ensures consistency with the prior. Conversely, for any $\gamma(l,l) \in (0,1]$ and $P > 0$ satisfying (C) there exists a feasible G.

The Fully Reduced Problem. By Lemma B.4 all breakdown probabilities are linear in $\gamma(l,l)$. If we plug all breakdown probabilities into the aggregate feasibility constraint, (AF), we get an expression of the form

$$\underbrace{2V(l) - \gamma(l,l)Q(P)}_{\text{LHS of }(AF)} \leq \underbrace{1 - \gamma(l,l)R(P)}_{1-Pr(\Gamma)}, \tag{B.1}$$

where $\gamma(l,l)Q(P) := \sum_i \sum_m p(m)z_i(m) - 2V(l)$. Assumption 1 implies $Q(P) \geq R(P)$ and we can reformulate

$$1 \geq \gamma(l,l) \geq \frac{\nu}{Q(P) - R(P)} =: \gamma^*(P), \tag{AF'}$$

with $\nu = 2V(l) - 1$ independent of P. Reducing $\gamma(l,l)$ reduces $Pr(\Gamma)$. Thus, constraint (AF') binds at the optimum, and $\gamma(l,l) = \gamma^*(P)$. Plugging into $Pr(\Gamma)$, we get

$$\min_P R(P)\gamma^*(P) \tag{P1$'$}$$

subject to the remaining constraints (IC_i^l), (IF), (EPI) and $\gamma^*(P) \leq 1$ and any solution to ($P1$) is also a solution to ($P1'$).[1]

Irrelevance of Signals.

Lemma B.5. *It is without loss of generality to assume that the mediator does not release public signals.*

Proof. Suppose we augment the mediator's protocol by public signals, being released after the breakdown of mediation. Let S be the set of these signals, with generic element s. If $\Gamma = 1$ denotes the event of breakdown, and $\Gamma = 0$ denotes the event of settlement, then $Pr(s, \Gamma = 1, c_i, c_{-i})$ is the joint probability that signal s and the type profile c_i, c_{-i} realizes and breakdown occurs. Each realized signal s gives rise to a system of first-order beliefs, i.e. player's beliefs about the other player's payoff-type. Thus, each such element gives rise to a different play of the litigation game. Denote by $U_i(c|c_{-i}, s, \hat{c})$ type c's payoff in this game, when having reported to be $\hat{c}$ and choosing an optimal strategy and his opponent is of type c_{-i}. Then, player i type h's incentive constraint reads:

$$\sum_{s\in S}\sum_{c_{-i}}[Pr(s,\Gamma=0,\hat{c}_i=h,c_{-i})z_i(s,\hat{c}_i=h,c_{-i}) + Pr(s,\Gamma=1,\hat{c}_i=h,c_{-i})U_i(h|c_{-i},s,\hat{h})]$$

$$\geq \sum_{s\in S}\sum_{c_{-i}}[Pr(s,\Gamma=0,\hat{c}=l,c_{-i})z_i(s,\hat{c}=l,c_{-i}) + Pr(s,\Gamma=1,\hat{c}_i=l,c_{-i})U_i(h|c_{-i},s,\hat{l})]$$

[1] Problem (P1′) is in fact equivalent to problem ($P1$) whenever $P > 0$. As every argument is continuous in P this limitation only becomes relevant once (P1′) has no minimum.

Let $z_i(\hat{c}) = \sum_{s\in S}\sum_{c_{-i}} Pr(s,\Gamma=0,\hat{c}_i,c_{-i})z(s,\hat{c}=h,c_{-i})$ and note

$$\sum_{c_{-i}} Pr(s,\Gamma=1,\hat{c}_i,c_{-i})U_i(c_i|c_{-i},s,\hat{c}_i) = Pr(s,\Gamma=1,\hat{c}_i)U_i(c_i|s,\hat{c}_i)$$

Hence, the incentive constraint becomes:

$$z_i(h) - z_i(l) \geq \sum_{s\in S} Pr(s,\Gamma=1,\hat{c}_i=l)U_i(h|s,\hat{l}) - Pr(s,\Gamma=1,\hat{c}_i=h)U_i(h|s,\hat{h})$$

Moreover, type l's participation constraint becomes:

$$z_i(l) + \sum_{s\in S} Pr(s,\Gamma=1,\hat{c}_i=l)U_i(l|s,\hat{l}) \geq V(l)$$

Given those constraints, let us repeat the steps from above. Then, $Q(P)$ and $R(P)$, being defined in equation (B.1), satisfy the following relation:

$$\frac{Q(P)}{R(P)} = \sum_i\sum_{s\in S}\{\frac{Pr(s,\Gamma=1,l)}{R(P)}[U_i(l|s,l)-(1-p)U_i(h|s,l)] + \frac{Pr(s,\Gamma=1,h)}{R(P)}(1-p)U_i(h|s,h)\}$$

Let us define $H_i(l,s) := \frac{1}{p}[U_i(l|s,l) + (U_i(l|s,l)-(1-p)U_i(h|s,l))]$ and $H_i(h,s) := U_i(h|s,h)$. Then

$$\frac{Q(P)}{R(P)} = \sum_i\sum_{s\in S}\sum_{c_i}\frac{Pr(s,\Gamma=1,c_i)}{R(P)}p(c_i)H_i(c_i,s)]$$

Now, observe that by Bayes' rule $\frac{Pr(s,\Gamma=1,c_i)}{R(P)}p(c_i) = Pr(s,c_i|\Gamma=1) = Pr(c_i|s,\Gamma=1)Pr(s|\Gamma=1)$. Thus:

$$\frac{Q(P)}{R(P)} = \sum_i\sum_{s\in S}\sum_{c_i} Pr(c_i|s,\Gamma=1)Pr(s|\Gamma=1)H_i(c_i,s)$$

Since $Pr(s|\Gamma=1)$ is independent of i and c_i we arrive at:

$$\frac{Q(P)}{R(P)} = \sum_{s\in S} Pr(s|\Gamma=1)\sum_i\sum_{c_i} Pr(c_i|s,\Gamma=1)H_i(c_i,s)$$

First, observe that $\frac{Q(P)}{R(P)}$ is maximized for degenerated signals. Denote by s^* the solution to $arg\max_{s\in S}\sum_i\sum_{c_i} Pr(c_i|s,\Gamma=1)H_i(c_i,s)$. $\frac{Q(P)}{R(P)}$ is therefore maximized whenever s^* is released with certainty. Finally note that any P that maximizes $\frac{Q(P)}{R(P)}$ also solves equation (P1'). The objective of the latter problem reads $v\frac{R(P)}{R(P)}\frac{1}{\frac{Q(P)}{R(P)}-1}$ and is minimized if and only if $\frac{Q(P)}{R(P)}$ is maximized.

Public signals improve the mechanism compared to the optimal solution without public signals if and only if only one low-cost type's incentive constraint is binding. In this case, the mediator can release two public signals s_1 and s_2, each of which induce the same posteriors, but with players' roles reversed. In this way, slack from one player's low-types incentive constraint can be used to ease the other player's low-types incentive constraint. Besides this, the solution has the same feature as the one presented in lemma 3.7. Hence, for finding the optimal solution, we can focus on the same problem as without public signals, but replace the low cost-types' incentive constraints by a pooled version. □

B.2 Forces of Asymmetry

We first consider the optimal symmetric mechanism. Notice that the designer of a symmetric mechanism has only one choice variable $\tilde{\rho} := \rho_1 = \rho_2$. In a symmetric mechanism, Corollary 3.1 holds and any subscripts can be dropped. In combination with type-independent beliefs we get $U(h|h) = U(h|l) = 0$. By incentive compatibility, (IC^h), settlement values must thus be equal, i.e. $z(l){=}z(h){=}z$. Using the participation constraint, (PC^l), the settlement value z can be expressed as

$$z = V(l) - \gamma(l)U(l|l).$$

Ignoring any effect on $U(l|l)$, an increase in $\tilde{\rho}$ increases the settlement-value the mediator needs to offer. This effect is strengthened as $\tilde{\rho}$ decreases $U(l|l)$. Next, consider the total resources distributed

$$2z = 1 - Pr(\Gamma). \tag{AF}$$

As $\tilde{\rho}$ increases, breakdown decreases and the mediator can distribute more resources in case of settlement.
Combining the two equations yields

$$\underbrace{2V(l) - 1}_{=\nu} = 2\gamma(l)U(l|l) - Pr(\Gamma). \tag{B.2}$$

Using Corollary 3.2 and (3.2) we can rewrite equation (B.2)

$$\begin{aligned} \nu =& \gamma(l)\left((1-\tilde{\rho})\frac{(\kappa-1)}{\kappa} - \frac{p}{\tilde{\rho}}\right) \\ \Leftrightarrow \nu =& 2\underbrace{\gamma(l)\frac{p}{\tilde{\rho}}}_{=Pr(\Gamma)}\left(\frac{(1-\tilde{\rho})\tilde{\rho}}{p}\frac{(\kappa-1)}{\kappa} - 1\right). \end{aligned}$$

Solving for $Pr(\Gamma)$ yields

$$Pr(\Gamma) = \frac{\nu}{2}\left(\frac{(1-\tilde{\rho})\tilde{\rho}}{p}\frac{(\kappa-1)}{\kappa} - 1\right)^{-1}$$

which is minimized for $\tilde{\rho} = 1/2$. Thus, the optimal symmetric solution to $(P1')$ is obtained for breakdown probability $\tilde{\rho} = 1/2$.
A symmetric mechanism is, however, never optimal. This follows from the differences in the resources needed to sustain a certain level of either ρ_i. First, observe that despite any asymmetry, (IC_2^h) still requires that the settlement value of the high type $z_2(h) = z_2(l)$. As $U_2(h|h) = 0$, the breakdown value is 0 and expected

settlement valuation of player 2 is

$$z_2 := z_2(l) = V(l) - \gamma_2(l)U(l|l) = V(l) - \frac{(1-\rho_2)}{\rho_1}\frac{(\kappa-1)}{\kappa}p\gamma(l,l).$$

The first equality comes from (PC_2^l) and second from the results of Corollary 3.2 and the equations in (3.2).

For player 1, on the other hand the results change more substantially under asymmetry. Player $1h$'s incentive constraint is

$$z_1(h) = z_1(l) + \big(\gamma_1(l) - \gamma_1(h)\big)U_1(h|h). \tag{IC_1^h}$$

As $U_1(h|h) \neq 0$ the mediator pays an information rent to player 1 if $\gamma_i(l) \neq \gamma_i(h)$. Thus, the ex-ante expected valuation of player 1 under settlement is

$$\begin{aligned} pz_1(l)+(1-p)z_1(h) &= z_1(l)+(1-p)\big(\gamma_1(l) - \gamma_1(h)\big)U_1(h|h) \\ &= z_1(l) + \gamma_1(l)\left(1 - \frac{p}{\rho_1}\right)U_1(h|h) \end{aligned} \tag{B.3}$$

where the first uses (IC_1^h) and the second uses Corollary 3.2 to simplify. Simplifying this using (PC_1^l), (3.2), and $U_i(\cdot,\cdot)$ yields

$$z_1 := \underbrace{V(l) - \left(\frac{(1-\rho_1)}{\rho_2}\right)\frac{\kappa-1}{\kappa}p\gamma(l,l)}_{\text{symmetric part}} + \underbrace{\left(\frac{p}{\rho_1} - \frac{p}{\rho_2}\right)\frac{\kappa-1}{\kappa}p\gamma(l,l)}_{\text{asymmetric part}}.$$

While the symmetric part is always present, the asymmetric is only non-zero in asymmetric cases. As $\rho_1 > \rho_2$ in such cases the asymmetric part is genereically negative. Marginal effects on the second part cancel out with those on z_2. As the asymmetric part is additive separable in ρ_i, the optimum of ρ_i is independent of the choice of ρ_{-i}.

B.3 Proofs

B.3.1 Proof of Lemma 3.1

Proof. The proof is along the lines of Siegel (2014). However, as the proof is instructive and our setup differs slightly, we spell it out here. We first show that at least one type of one player has 0 expected utility. Second, we show that at most one type has an atom at 0. Third, we constructively show that the equilibrium exists and then show that it is indeed unique given (M). Then we calculate Δ to state Corollary 3.1.

Step 1: One player has 0 expected utility and no atoms at positive scores. We prove this by contradiction. Suppose that both players and both types expect a utility larger 0. That means the smallest score $\underline{s} > 0$ in the union of the best-responses of all players wins the contest with positive probability as otherwise it is no best response. As a result, the smallest score is an atom in the strategy of at least one type of each player. But then, there exists an ϵ in the neighborhood of $\underline{s}$ such that the probability of winning increases with more than $\epsilon * \kappa c_l$. Deviating to $\underline{s} + \epsilon$ is profitable for that type of player, and thus $\underline{s}$ cannot be an atom in her strategy. Therefore, at least one player earns an expected utility of 0 for sure. Note that this player may very well have an atom at 0 as there is no need to win the good with positive probability for an atom at 0. However, if both players had a type with an atom at 0 at least one of them can profitably deviate to a positive neighborhood of 0 winning against the atom scoring opponent with a probability that exceeds the cost of scoring. Thus, at most one player has an atom at 0.

Step 2a: Construct the equilibrium. First, consider the following strategy of player $2l$: she uniformly mixes on $(\Delta_{l,h}, \Delta_{l,l}]$ with density $f_{2,l}(t) = c_l/p_2(l,l)$. Then, player $1l$ is indifferent between playing any point on $s \in (\Delta_{l,h}, \Delta_{l,l}]$ as

$$\begin{aligned} U_1(l,s) &= F_2(\Delta_{l,h}) + p_2(l|l)(s-\Delta_{l,h})\frac{c_l}{p_2(l|l)} - c_l s = \\ &= F_2(\Delta_{l,h}) - \Delta_{l,h}c_l. \end{aligned}$$

We want to construct strategies with constant density and non-overlapping strategies, thus the length of the top interval $L(t)$ is the solution to

$$L(t)f_{2,l}(t) = 1.$$

To make player $2l$ indifferent as well, player $1l$ plays a similar strategy only flipping the probabilities from p_1 to p_2. As we assumed $p_1(l|l) \geq p_2(l|l)$, the mass of player $1l$ is only fully exhausted on the top interval iff $p_1(l|l) = p_2(l|l)$. If this is not the case, player 1 has some mass left to place. She does so on the middle interval $(\Delta_{h,h}, \Delta_{l,h}]$. For the same reasons as above, she assigns density $f_{1,l}(t) = c_l/p_1(l|h)$ to this interval to make player $1h$ indifferent.

The length of the medium interval can be calculated by acknowledging that player $1l$ needs to place all mass available to her and not placed on the top interval on this interval.

By a similar exercise we can find the length of the interval $(0, \Delta_{h,h})$ and by this the absolute values of all Δ.

Step 2b: Show that no (global) deviation is possible. What remains to be shown is that any player that scoring on more than one interval is in fact indifferent between those and that no global deviation is possible.

Note that the indifference across intervals follows from the intervals being connected. Consider for example player $1l$. From the above we know that

$$U_1(l, s = \Delta_{l,h}) = U_1(l, s = \Delta_{l,l})$$

but also that

$$U_1(l, s = \Delta_{h,h}) = U_1(l, s = \Delta_{l,h}).$$

Thus, it must be the case that

$$U_1(l, s = \Delta_{h,h}) = U_1(l, s = \Delta_{l,l}).$$

The same holds true for player $2h$. The two other player-type tuples place there scores on a single interval only. Note that, since player $1h$ has positive mass only on $(0, \Delta_{h,h}]$ it can in fact earn an expected utility greater 0 if and only if player $2h$ does not enter the auction with positive probability.

To exclude global deviation observe that player $2h$ would only deviate to anything on the interval $(\Delta_{l,h}, \Delta_{l,l}]$ if the probability of winning increases faster in the top interval than in the middle interval, that is the density is smaller in the top interval,

$$f_{1,l}(m) = \frac{c_l}{p_1(l|l)} \geq \frac{\kappa c_l}{p_1(l|h)} = f_{1,l}(t),$$

which is ruled out by (M).
For $1h$, the deviation could be made into the middle or the top interval if

$$\frac{\kappa c_l}{p_2(h|h)} \geq \frac{c_l}{p_2(h|l)},$$

which again is ruled out by monotonicity. As player $1h$ prefers the bottom interval to anything in the m she must prefer scoring at $\Delta_{l,l}$ to $\Delta_{h,l}$. However as player $2h$ does not prefer to score at $\Delta_{l,l}$ it follows that $\Delta_{l,l} > 1/\kappa c_l$. Thus player $1h$ does not want to deviate. Similar arguments hold for the second player, such that we can conclude that global deviations are not beneficial.
Step 3: Uniqueness. For uniqueness observe first that there is only one monotonic equilibrium, that is an equilibrium such that the lowest score of player i, type l, is weakly above the highest score of player i, type h. This follows directly from the equilibrium construction.
Second, we need to show that no non-monotonic equilibrium exists. We do so by contradiction, that is suppose there exists a score $s_i^h > s_i^l$ such that s_i^k is in the set of best responses for player i type k, $BR(k)$. Then, it must hold that

$$\begin{aligned} & U_i(h, s = s_i^h) \geq U_i(h, s = s_i^l) \\ \Leftrightarrow \quad & \sum_k p_i(k|h) F_{-i,k}(s_i^h) - \kappa c_l s_i^h \geq \sum_k p_i(k|h) F_{-i,k}(s_i^l) - \kappa c_l s_i^l \\ \Leftrightarrow \quad & \sum_k p_i(k|h)(F_{-i,k}(s_i^h) - F_{-i,k}(s_i^l)) \geq \kappa c_l (s_i^h - s_i^l). \end{aligned} \tag{B.4}$$

Similarly, as s_i^l is a best response for l it must hold that

$$\sum_k p_i(k|l)(F_{-i,k}(s_i^h) - F_{-i,k}(s_i^l)) \leq c_l (s_i^h - s_i^l). \tag{B.5}$$

But, as $F_{-i,k}(\cdot)$ is always positive and $p_i(h|\cdot) = 1 - p_i(l|\cdot)$, inequalities (B.4) and (B.5) only hold if

$$\frac{p_i(l|h)}{\kappa c_l} \sum_k (F_{-i,k}(s_i^h) - F_{-i,h}(s_i^l)) \geq \frac{p_i(l|l)}{c_L} \sum_k (F_{-i,k}(s_i^h) - F_{-i,h}(s_i^l)).$$

As the sum is identical on both sides, this boils down to the inverse of (M), a contradiction. □

(Addendum) Step 4: Equilibrium expected utilities. The length of the top interval, $(\Delta_{l,h}, \Delta_{l,l}]$, is $p_2(l|l)/c_l$ that of the bottom interval, $(0, \Delta_{h,h})$, is $p_1(h|h)/\kappa c_l$ and that of the middle interval $(\Delta_{h,h}, \Delta_{l,h}]$ is

$$\frac{p_1(l|h)}{\kappa c_l}(1 - \frac{f_{1,l}(t)}{f_{2,l}(t)}) = \frac{p_1(l|h)}{\kappa c_l}(1 - \frac{p_2(l|l)}{p_1(l|l)}).$$

Putting the respective probability masses on the different intervals leaves player 2 with some mass $\mu \geq 0$. This is placed on scoring 0 and constitutes $F_{2,h}(0)$.

Notice that scoring $\Delta_{l,l}$ wins the auction for sure at cost of $\Delta_{l,l}c_l$ for both players, type l, and player 1 scoring (arbitrarily close) to 0 wins the auction with probability $F_{2,h}(0)$ at almost no cost.

B.3.2 Proof of Lemma 3.2

Proof. First, consider player $2h$. She earns an expected utility of 0 on-path. Post-deviation she can always choose a score of 0 to secure this utility.

Second, consider player il. Independently of her report she can always choose a score $\Delta_{l,l}$ and win with probability 1. As this is also part of the best response on-path and the probability is 1 in that case as well, she can only be better of by choosing a score different than $\Delta_{l,l}$.

Finally consider player $1h$ after reporting to be type l. She holds belief $p_2(h|l)$ while her opponent plays the equilibrium strategies. If she were to score 0, then by our tie-braking assumption she would enjoy a utility at least as good as the equilibrium utility if $p_2(h|l) \geq p_2(h|h)$. Thus, in those cases she is weakly better of.

If, however, $p_2(h|l) < p_2(h|h)$ then player 1 suffers whenever scoring against an h-type compared to the on-path game as the probability of winning decreases while costs stay the same. However, scoring against the low-cost type and at the same time earning a higher expected utility than in the default game can, by the constant density of player 2's low-cost type on the support of her equilibrium strategy, only mean scoring to the very top, that is $\Delta_{l,l}$ which yields negative utility to a high type by the construction of the equilibrium. □

B.3.3 Proof of Lemma 3.3

Proof. First, notice that player il benefits if and only if $p_{-i}(l|l) > p_{-i}(l|h)$. The if part follows directly from the density of the opposing player on the top interval which is $f_{-i,l}(t) = c_l/p_{-i}(l|l)$. As $p_{-i}(l|h)$ is smaller than this, scoring at $\Delta_{l,h}$ is strictly preferred to $\Delta_{l,l}$, but $\Delta_{l,l}$ yields the same result as the on-path game.

The only-if party follows as for $p_{-i}(l|l) = p_{-i}(l|h)$ would induce type independent beliefs and therefore the same result as the on-path game. For $p_{-i}(l|l) < p_{-i}(l|h)$, however, scoring at the top, i.e. $\Delta_{l,l}$ is preferred leading to no changes in expected utilities at all.

As $p_{-i}(l|l) < p_{-i}(l|h)$ implies $p_{-i}(h|h) < p_{-i}(h|l)$ we know that player 1, type h is better off, as scoring 0 yields him already a higher payoff by $p_2(h|l)F_{2,h}(0) > p_2(h|h)F_{2,h}(0)$. Player 2 strictly prefers to score at $\Delta_{h,h}$ compared to 0 as the density of her opponent is given by $f_{1,l}(b) = c_l/p_1(h|h)$ which leads to a (strictly) increasing utility on the bottom interval. Thus scoring at $\Delta_{h,h}$ must yield strictly positive utility.

The only setup in which neither party has a type that strictly profits from deviating is that of type-independent beliefs. □

B.3.4 Proof of Proposition 3.1 (together with lemmas B.1 to B.4)

The proof of the proposition is along the lines described in appendix B.1.

Proof of Lemma B.1

Proof. We show that the condition stated on the optimality of full participation stated in Proposition 2 of Celik and Peters (2011) is satisfied. That is, given the independent prior p, there is no Bayes' plausible belief structure $\tilde{p} = (\underline{p}, \overline{p})$ such that the expected utility $U_i(k, \tilde{p}, p) < U_i(k, p, p)$ for any type k. The condition is a direct consequence of expected contest utilities under a type-independent prior as defined in Corollary 3.1. For type independent priors utilities are in fact linear in beliefs except for a kink at the point where utilities become flat. However, around that point utilities are convex and Jensen's inequality yields the desired result. □

Proof of Lemma B.2

Proof. We apply theorem 3 of Border (2007) which says the following:

> **Border (2007), Theorem 3:** The list **P**= $(P_1, ..., P_N)$ of functions is the reduced form of a general auction **p**= $(p_1, ...p_n)$ if and only if for every subset $A \subset \mathcal{T}$ of individual-type pairs (i, τ) we have
>
> $$\sum_{(i,\tau)\in A} P_i(\tau)\mu^\bullet(\tau) \leq (\{t \in T : \exists (i,\tau) \in A, t_i = \tau\}).$$

An individual type pair in our setting is given by (m, i), in what follows we are going to abuse notation slightly by treating $p(m)$ such that $p(l) = p$ and $p(h) = 1-p$. The general auction **p** in our setup is defined by a list

$$q_i(m, n) := x_i(m, n).$$

We want to implement **p** by the list **P** containing

$$Q_i(m) := q_i(m, l)\mu_i(l|m) + q_i(m, h)\mu_i(h|m)$$

where

$$\mu_i(n|m) := \frac{\mu(m, n)}{\mu_i^\bullet(m)},$$
$$\mu(m, n) := p(l)p(m)\frac{1 - \gamma(m, n)}{1 - Pr(\Gamma)},$$
$$\mu_i^\bullet(m) := p(m)\frac{1 - \gamma_i(m)}{1 - Pr(\Gamma)}.$$

Plugging in yields,

$$Q_i(m) = \frac{p(l)(1 - \gamma(m, l))x_i(m, l) + p(h)(1 - \gamma(m, h))x_i(m, h)}{1 - \gamma_i(m)} = x_i(m).$$

To state the conditions let in addition

$$m^c := \{y \in \{l, h\}|y \neq m\}.$$

Applying the above quoted theorem of Border (2007) to this and reformulating everything in terms of z_i allows us to conclude that $\mathcal{X}$ can be implemented via $z_i \geq 0$ if and only if the following conditions are satisfied:

- $\forall\{m,n\} \in \{h,l\}$:
$$p(m)z_i(m) + p(n)z_{-i}(n) \leq \qquad (EPI)$$
$$1 - Pr(\Gamma) - (1-\gamma(m^c,n^c))p(m^c)p(n^c)$$

- $\forall m \in \{h,l\}$ and $i = 1,2$:
$$z_i(m) \leq 1 - \gamma_i(n) \qquad (IF)$$

- $\forall i = 1,2$
$$z_i(l)p(l) + z_i(h)p(h) \leq 1 - Pr(\Gamma) \qquad (BC_2)$$
$$\sum_{i\in\{1,2\}} \sum_{k\in\{l,k\}} p(k)z_i(k) \leq 1 - Pr(\Gamma) \qquad (AF)$$

- $\forall\{m,n\} \in \{h,l\}^2$ and $i = 1,2$:
$$\sum_{k\in\{l,h\}} p_i(k)z_i(k) + pz_{-i}(n) \leq 1 - Pr(\Gamma). \qquad (BC_4)$$

Note that in our setup equation (IF) implies (BC_2) and equation (AF) which implies (BC_4). For the second claim, recall (IC_i^l), that is

$$\gamma_i(h)U_i(h|l) + z_i(h) \leq \gamma_i(l)U_i(l|l) + z_i(l).$$

Hence,

$$z_i(h) \leq \gamma_i(l)U_i(l|l) - \gamma_i(h)U_i(l|h) + z_i(l) \leq (\gamma_i(l) - \gamma_i(h))U_i(l|l) + z_i(l), \qquad \text{(B.6)}$$

where the last equality follows from Lemma 3.2.
If $\gamma_i(l) \geq \gamma_i(h)$ and $z_i(l) \leq 1 - \gamma_i(l)$ we can rewrite (B.6) to

$$z_i(h) \leq (\gamma_i(l) - \gamma_i(h))U_i(l|l) + z_i(l) \leq 1 - \gamma_i(h),$$

which indeed is equation (IF). □

Proof of Lemma B.3

Proof. We proof this by contradiction. Suppose there exists a feasible $\mathcal{X}$ that forms an optimal mediation protocol without (IC_i^h) binding for some i. That is, without loss of generality assume that for player 1 it holds that

$$z_1(h) - z_1(l) > \gamma_1(h)U_1(h|h) - \gamma_1(l)U_1(h|l).$$

Recall that

$$z_1(h) = p(1-\gamma(h,l))x_1(h,l) + (1-p)\gamma(h,h)x_1(h,h),$$

but then if $\mathcal{X}$ was feasible before, it remains feasible if we reduce $x_1(h,l)$ such that (IC_1^h) holds with equality. Changing this has no effect on the right hand side of the inequality and (IC_i^l) gets relaxed as it is

$$z_1(h) - z_1(l) \leq \gamma_1(h)U_1(l|h) - \gamma_1(l)U_1(l|l)$$

Similarly, suppose (PC_i^l) is not binding, then

$$z_i(l) > V_i(l) - \gamma_i(l)U_i(l|l).$$

Provided that $z_i(l) > 0$ the mediator could react, by changing $z_i(l)$ such that the participation constraint is binding. Then, she can reduce $z_1(h)$ such that the high-cost types incentive constraint is binding which leads to another $\mathcal{X}$ with both (PC_i^l) and (IC_i^h) binding that is feasible and delivers the same value to the objective.

If $z_i(l) = 0$, this procedure is not possible, but then the mediator could use the homogeneity of degree 1 of $\gamma_i(k)$ and the homogeneity of degree 0 w.r.t. G to satisfy (PC_i^l) by multiplying all elements of G by $\alpha < 1$. Again, if $z_i(h) > 0$ the increase in in $z_i(h)$ can always be off-set by reducing X_i appropriately which is always possible. If $z_i(h)$ is indeed 0, then multiplying G by α has if at all only a positive effect on incentive compatibility. Thus, it is without loss of generality to assume that (PC_i^l) holds indeed. □

Proof of Lemma B.4

Proof. Recall that the elements of P can be rewritten such that e.g. the probability of meeting player $1l$, given a report $m_2 = l$ is

$$p_1(l|l) = \frac{p\gamma(l,l)}{p\gamma(l,l) + (1-p)\gamma(h,l)}. \tag{B.7}$$

As $p_1(l|l) > 0$ which is guaranteed by $\gamma(l,l) > 0$ the probability representation for $\gamma(h,l)$ follows immediately, that is

$$\gamma(h,l) = \frac{1 - p_1(l|l)}{p_1(l|l)} \frac{p}{1-p} \gamma(l,l).$$

Repeating the same exercise for any $\gamma(k,m)$ yields the desired representation.

The last equation of (C) can be obtained noticing that given we have established all other results from (C) and using the homogeneity of degree 0 of P w.r.t G we can rewrite G as

$$G = \gamma(l,l)G' = \gamma(l,l) \begin{pmatrix} 1 & q_2(l) \\ q_1(l) & q_2(h)q_1(l) \end{pmatrix}.$$

We know that G' induces the same P as G in particular we know that

$$p_1(l|h) = \frac{p\gamma(h,l)}{p\gamma(h,l) + (1-p)\gamma(h,h)} = \frac{pq_2(l)}{pq_2(l) + (1-p)q_2(h)q_1(l)}$$

which after rearranging yields the desired

$$q_1(l)q_2(h) = q_1(h)q_2(l). \tag{C}$$

As all we have done have been rearrangements, the converse holds as well, that is, for a given P and $\gamma(l,l) > 0$ that satisfy equation (C) we can establish a feasible G such that P and $\gamma(l,l)$ is induced by G. $\square$

B.3.5 Proof of Theorem 3.1 (together with Lemmas 3.4 to 3.7)

We proof the proposition in several steps. In line with the text, we first solve the "unconstrained problem" $(P1')$ which is also the proof of Lemma 3.5.[2] After that we introduce (IC_i^l) and proof Lemma 3.6 before finally introducing the remaining constraints with the proof of Lemma 3.7. Throughout this proof we make use of the following lemma

Lemma B.6. *At any optimum of* $(P1')$*, the monotonicity condition* (M) *is always satisfied.*

The proof of this lemma can be found at the end of the appendix as it is neither constructive nor relevant to understand the main argument. However, with help of this lemma, we can restrict the choice set of the mediator to the set of induced beliefs that result in monotonic equilibria as discussed in Lemma 3.1.

Proof of Lemma 3.4

Proof. Rewrite $p_2(l|h)$ with help of Lemma B.4

$$p_2(l|h) = \frac{\Big(1 - p_1(l|l)\Big)\, p_2(l|l)p_1(l|h)}{p_2(l|l)p_1(l|h) - p_1(l|l)\Big(1 - p_2(l|l) - p_1(l|h)\Big)}. \tag{B.8}$$

$p_2(l|h) > p_2(l|l)$ if equation (B.8) divided by $p_2(l|l)$ is larger 1 that is

$$\frac{p_2(l|h)}{p_2(l|l)} = (1 - p_1(l|l))\frac{p_1(l|h)}{p_2(l|l)p_1(l|h) - p_1(l|l)\Big(1 - p_2(l|l) - p_1(l|h)\Big)} > 1.$$

Rewriting yields,

$$\begin{aligned} & (1 - p_1(l|l))p_1(l|h) > p_2(l|l)p_1(l|h) - p_1(l|l)(1 - p_2(l|l) - p_1(l|h)) \\ \Leftrightarrow \quad & p_1(l|h) - p_1(l|l) > (p_1(l|h) - p_1(l|l))p_2(l|l), \end{aligned}$$

which holds if and only if $p_1(l|h) > p_1(l|l)$. $\square$

Proof of Lemma 3.5

Proof. Notice that the unconstrained problem is $(P1')$ is a problem of three elements $P = (p_1(l|l), p_2(l|l), p_1(l|h)$ only, as the fourth is directly defined via consistency equation (C). We calculate the unconstrained optimum in several steps. First, we show that at the optimum the objective is not differentiable with respect to at least one of the three choices variables. Second, we show that if $p_1(l|l)$ is either $p_1(l|h)$ or $p_2(l|l)$m then it is $p_1(l|l) = p_1(l|h)$ and calculate this optimum. Finally, we show that a deviation to $p_1(l|l) = 1$ is not optimal.

[2]Recall that "unconstrained" refers to $(P1')$ which includes all constraints that bind at all points already in the problem definition.

Step 1: No optimum in the differentiable interior exists. To proof this claim we are going to proof that the objective $Obj(P) := R(P)v/(Q(P) - R(P))$ is locally concave at any critical point in $p_1(l|l)$ in what we call the "differentiable interior", meaning that such a critical point is in fact a local maximum in $p_1(l|l)$, which is sufficient to proof the claim. Let us begin with defining the differentiable interior.

Definition B.1 (Differentiable Interior)**.** The differentiable interior of problem (P1′) is the set of all P such that for each $1 > p(k|m) > 0$ the left-derivative and the right-derivative of $Obj(P)$ with respect to all variables coincides.

Next, for the ease of notation define $\boldsymbol{\rho} = (\rho_1(l), \rho_2(l), \rho_1(h)) := (p_1^*(l|l), p_2^*(l|l), p_1^*(l|h))$
Step 1a: Transform $R(P)$ and $Q(P)$.
Observe that

$$R(P) = \frac{Pr(\Gamma)}{\gamma(l,l)} = \frac{p^2}{\rho_1(l)\rho_2(l)\rho_1(h)} \left(\rho_1(l)(1 - \rho_2(l)) + \rho_2(l)\rho_1(h)\right).$$

Defining the function

$$\tilde{Y} := Y * \frac{\rho_1(l)\rho_2(l)\rho_1(h)}{p^2}$$

allows us to rewrite (dropping the argument to simplify notation)

$$Obj = \frac{\tilde{R}v}{\tilde{Q} - \tilde{R}}.$$

Notice that $\tilde{R}$ is linear in any variable of $\boldsymbol{\rho}$. **Step 1b: Define necessary conditions for an optimal interior point.** Suppose $(P1')$ has indeed an optimal point in the differentiable interior. Then a necessary condition on this point is that it is indeed a critical point in all three variables, that is

$$Obj'(\rho) := \frac{\partial Obj(\rho)}{\partial \rho} = \underbrace{\frac{\nu}{(\tilde{Q}(\rho) - \tilde{R}(\rho))^2}}_{=:f(\rho)} \underbrace{\left(\tilde{R}'(\rho)\tilde{Q}(\rho) - \tilde{Q}'(\rho)\tilde{R}(\rho)\right)}_{=:g(\rho)} = 0 \qquad \text{(FOC)}$$

for every $\rho \in \boldsymbol{\rho}$. Noticing that $f(\rho) \neq 0$ for any ρ by definition, the necessary first order condition boils down to $g(\rho) = 0$. Another necessary condition for a local minimum is that any critical point in any ρ is not locally concave in this variable. If it was locally concave in any ρ this means that we are at a local maximum in this variable ρ and that the second order conditions for a minimum are never fulfilled. Formally, this means that at any critical point ρ^{cp} it needs to hold that

$$Obj''(\rho^{cp}) = \underbrace{f'(\rho^{cp})g(\rho^{cp})}_{=0 \text{ by equation (FOC)}} + f(\rho^{cp})g'(\rho^{cp}) \geq 0 \qquad \text{(B.9)}$$

for every $\rho^{cp} \in \boldsymbol{\rho^{cp}}$. The first term is 0 by the standard envelope argument, such that (B.9) boils down to

$$Obj''(\rho^{cp}) = f(\rho^{cp})g'(\rho^{cp}) = f(\rho^{cp}) \left(\tilde{R}''(\rho^{cp})\tilde{Q}(\rho^{cp}) - \tilde{R}(\rho^{cp})\tilde{Q}''(\rho^{cp})\right) \geq 0.$$

By the linearity of $\tilde{R}$ and the observation that $\tilde{R} \geq 0$ by construction, a necessary and sufficient condition for (B.9) to hold is simply

$$\tilde{Q}''(\rho^{cp}) \leq 0 \tag{B.10}$$

for every $\rho^{cp} \in \boldsymbol{\rho^{cp}}$. **Step 1c: Show that the necessary conditions never hold for $\rho_1(l)$.** To complete the claim of step 1 we are now going to show, that $\tilde{Q}(\rho_1(l))$ is indeed a convex function.

To see this observe first by plugging in we can reduce $\gamma_i(l) = \gamma(l,l)p/\rho_{-i}(l)$ which in turn means that while $\tilde{\gamma}_2(l)$ is constant in $\rho_1(l)$, $\tilde{\gamma}_1(l)$ is linearly increasing in $\rho_1(l)$. In addition, we do not need to worry about $\gamma_2(h)$ as player $2h$ has no expected utility by Corollary 3.1. Further we can rewrite using Corollary 3.1 and Lemma B.4

$$\tilde{\gamma}_1(h)U_1(h|h) = \frac{\gamma(l,l)}{1-p}(1-\rho_1(h))\rho_1(h)\ (\rho_1(l)-\rho_2(l))\frac{(\kappa-1)}{\kappa}$$

which is linearly increasing in $\rho_1(l)$ and positive. Rewriting yields

$$\gamma(l,l)\tilde{Q} = \sum_i \tilde{\gamma}_i(l)\left(U_i(l|l)-(1-p)U_i(h|l)\right) + \tilde{\gamma}_1(h)(1-p)U_1(h|h)$$

it suffices to show that

$$h_i(\rho_1(l)) = \tilde{\gamma}_i(l)\big(U_i(l|l)-(1-p)U_i(h|l)\big)$$

is convex for every i.

For h_2, observe that by Lemmas 3.2 and 3.3, player $2h$ only gains from deviating if $\rho_1(h) > \rho_1(l)$. In such a case player $2h$, best post-deviation strategy is to play $\Delta_{h,h}$ with probability 1, which yields utility

$$U_2(h|l) = p_1(h|l) - \Delta_{h,h}\kappa c_l. \tag{B.11}$$

Bidding the same on-path is in the best response set of player 2 yielding

$$U_2(h|h) = p_1(h|h) - \Delta_{h,h}\kappa c_l = 0. \tag{B.12}$$

Subtracting equation (B.12) from equation (B.11) yields

$$p_1(h|l) - p_1(h|h) = \rho_1(h) - \rho_1(l) = U_2(h|l) \tag{B.13}$$

and thus $U_2(h|l)$ is linear in $\rho_1(l)$. As $\tilde{\gamma}_2(l)$ is constant in $\rho_1(l)$, $h_2(\rho_1(l))$ is convex if and only if $U_2(l|l)$ is convex in $\rho_1(l)$ which can easily be verified by the utilities derived in (U). The last step is now to show that $h_1(\rho_1(l))$ is convex as well.

To see this, observe first that whenever deviation is profitable for player 1, type h, she would deviate by playing $\Delta_{h,h}$. But, $\Delta_{h,h}$ is in fact the lower bound of player 1, type l and thus in such a case we can rewrite

$$U_1(h|l, p_2(h|l) > p_2(h|h)) = U_1(l|l) + (1-\kappa)c_l\Delta_{h,h}.$$

As $\tilde{\gamma}_1(l) = \rho_1(l)\rho_1(h)p$ we can use the expression derived in Corollary 3.1 to establish that $\tilde{\gamma}_1(l)U_1(l|l)$ is linear in $\rho_1(l)$ and thus convex.

What remains is to show that $-\rho_1(l)\Delta_{h,h}$ is weakly convex. This can be established using that $\Delta_{h,h} = p_1(h|h)/\kappa c_l$ which is independent of $\rho_1(l)$ which proofs the claim.
Step 2: $\rho_1(l) \in \{\rho_2(l), \rho_1(h)\}$**.**
By assumption $\rho_1(l) \leq \rho_2(l)$ is ruled out. Second, fix some $\rho_2(l)$ and $\rho_1(h)$. If $\rho_1(l) \in [\rho_2(l), \rho_1(h)]$ then $Obj(\rho_1(l) = 1) > Obj(\rho_1(l) = \rho_1(h))$. Further we know that Obj is continuously differentiable on $\rho_1(l) \in (\max\{\rho_2(l), \rho_1(l)\}, 1)$. By Step 1 we know that every interior point is a maximum in $\rho_1(l)$.
Next, notice by Lemma 3.4 that for $\rho_1(l) > \rho_1(h) \Rightarrow \rho_2(l) > \rho_2(h) \Rightarrow p_i(h|h) > p_i(h|l) \Rightarrow U_2(h|l) = 0$ and $U_1(h|l) < U_1(h|h)$.
Now, notice that $\rho_1(l) = 1$ can only be optimal if Obj is (LHS-)decreasing at $\rho_1(l) = 1$ as there cannot be a local minimum in $\rho_1(l)$ by Step 1. To check this it suffices to look at the sign determining function of the derivative which is, by Step 1, $R'Q - Q'R$. Solving this for $\rho_1(l) > \rho_1(h)$ yields a quadratic function in $\rho_1(l)$.
The sign-determining function at $\rho_1(l) = 1$ is quadratic in $\rho_1(h)$, i.e. a condition

$$a\rho_1(h)^2 + b\rho_1(h) + c < 0 \tag{B.14}$$

where

$$a = (\kappa - 1 + \rho_2(l)^2) \tag{B.15}$$
$$b = 1 + 2\rho_2(l) - 2\big(\rho_2(l)\big)^2 + p(1-\kappa) \tag{B.16}$$
$$c = \big(\rho_2(l)\big)^2\kappa - \rho_2(l)\big((\kappa-1)(1-p) + \kappa\big) + (\kappa-1)(1-p). \tag{B.17}$$

Note first, that (B.14) is decreasing in $\rho_2(l)$, second note that for $\rho_2(l) = \rho_1(h)$ condition (B.14) becomes

$$(\kappa-1)(1-p-2\rho_1(h)) + \big(\rho_1(h)\big)^4 - 2\big(\rho_1(h)\big)^3 + \big(\rho_1(h)\big)^2(1+2\kappa) < 0. \tag{B.18}$$

Note that this is minimal if κ is minimal and p is maximal. Therefore, it must hold that

$$\underbrace{1/2 + \rho_1(h))^2 \underbrace{\Big((\rho_1(h))^2 - 2\rho_1(h) + 5\Big)}_{>4} - 2\rho_1(h)}_{>-1/4} < 0,$$

a contradiction. Thus, whenever $\rho_1(h) \geq \rho_2(l)$, choosing $\rho_1(l) = 1$ is not preferred to $\rho_1(l) = \rho_1(h)$. Solving the first order conditions given $\rho_1(l) = 1$ for $0 < \rho_1(h) < \rho_2(l)$ yields that no critical point in both variables exists and therefore no interior solution. As Obj is decreasing at $\rho_2(h) = 0$, there cannot be any solution with $\rho_1(l) = 1$. Thus, $\rho_1(l)$ must either be equal to $\rho_1(h)$ or to $\rho_2(l)$.
Step 3: Calculate the optimum if $\rho_1(l) \in \{\rho_2(l), \rho_1(h)\}$**.** By Step 1, we know that if $\rho_1(l) \in [\rho_2(l), \rho_1(h)]$ the optimum involves ρ_1 being equal to either of the bounds. Therefore, we only need to consider the two cases for any $\rho_2(l)$ and $\rho_1(h)$.
Step 3a: The equilibrium for $\rho_1(l) = \rho_2(l)$**.** First, consider $\rho(l) = \rho_1(l) = \rho_2(l)$. By Lemma B.4, $\rho_1(h) = \rho_2(h) = \rho(h)$.
All payoffs are symmetric and, by Corollary 3.1, $U_i(h|h) = 0$ and, by Lemma 3.2, $U_i(h|l) = \max\{0, \rho(h) - \rho(l)\}$. Finally, $U_i(l|l) = (\kappa-1)/\kappa + (\rho(h) - \rho(l)\kappa)/\kappa$.

In addition, $\gamma_1(l) = \gamma_1(h) = \gamma_2(l) = p/\rho(l)$ and therefore

$$\tilde{Q} = \frac{2\rho(l)\rho(h)}{p}(U_i(l|l) - (1-p)U_i(h|l)).$$

Finally, as $\tilde{R} = \rho(l)(1 - \rho(l) + \rho(h))$ we can simplify Obj to

$$Obj(\rho(l), \rho(h)) = \frac{p(1-\rho(l)+\rho(h))}{\underbrace{2\rho(h)(U_i(l|l) - (1-p)U_i(h|l))}_{=:\hat{Q}} - \underbrace{p(1-\rho(l)+\rho(h))}_{=:\hat{R}}}.$$

Employing the same technique as in Step 1, we know, as $\hat{R}$ is linear in both $\rho(k)$ any interior solution needs to have that $\hat{Q}$ is concave in $\rho(k)$.

Notice that the second derivative of $\hat{Q}$ when $U_i(h|l) = 0$ boils down to $4/\kappa$ as $U_i(l|l)$ is linearly increasing with factor $1/\kappa$ in $\rho(h)$. Thus, any solution with $\rho(l) \geq \rho(h)$ can be ruled out.

Second whenever $\rho(l) < \rho(h)$ observe that $\hat{Q}$ is linearly decreasing in $\rho(l)$ with factor $2\rho(h)p$. Hence, the sign determining function of the first derivative $\hat{R}'\hat{Q} - \hat{Q}'\hat{R}$ becomes

$$\hat{R}'(\rho(l))\hat{Q} - \hat{Q}'(\hat{(l)})\hat{R}|_{\rho(l)<\rho(h)} = -2\rho(h)p\Big((U_i(l|l) - (1-p)U_i(h|l)) - \hat{R}\Big). \quad \text{(B.19)}$$

Note that by construction Obj defines a probability and is thus in $[0, 1]$. Whenever equation (B.19)=0, then $\hat{Q} - \hat{R} = (2\rho(h)p - 1)\hat{R}$ which can only be positive if $2\rho(h)p = 1$. As $p < 1/2$ this condition never holds. Therefore, we do not find an interior solution when $\rho(h) > \rho(l)$.

What remains are then boundary solutions with either of the $\rho(k) \in \{0, 1\}$.

If $\rho(h) = 1$ we need to go back to the original Q and R as our modifications are not valid if $\rho_i(k) \neq (0, 1)$.

This is for $\rho(h) = 1$

$$R = p^2 \frac{2 - \rho(l)}{\rho(l)}$$

$$Q = p^2 \frac{2(1 - \rho(l))}{\rho(l)},$$

which obviously violates $Q > R$ and is thus not feasible. $\rho(l) = 0$ would violate monotonicity and is ruled out by Lemma B.6.

Step 3b: The equilibrium for $\rho_1(l) = \rho_1(h)$. It remains to show that an equilibrium exists in which $\rho_1(l) = \rho_1(h) = \rho_1$. Note that again by consistency in Lemma B.4 we get $\rho_2 = \rho_2(l) = \rho_2(h)$.

With this, we know that $U_i(k|m) = U_i(k|k)$ for every i and k and $U_2(h|l) = 0$, $U_i(l|l) = (1 - \rho_2)\frac{\kappa - 1}{\kappa}$, and $U_1(h|h) = U_1(h|l) = (\rho_1 - \rho_2)\frac{\kappa - 1}{\kappa}$. As $\tilde{\gamma}_i(l) = p/\rho_{-i}$ we get

$$\tilde{Q} = \frac{1}{\kappa p}\rho_1(\kappa - 1)((\rho_1)^2 - \rho_1(1 + p) - \rho_2(1 - \rho_2 - p))$$

and

$$\tilde{R} = \rho_1.$$

Note that this means that for an optimum in ρ_1 we need $\tilde{Q} = \rho_1 \tilde{Q}'(\rho_1)$ and for an optimum in ρ_2 we would need $\tilde{Q}'(\rho_2) = 0$.
Notice that

$$\tilde{Q}'(\rho_2) = \frac{\rho_1(\kappa - 1)}{\kappa p}(1 - p - 2\rho_2)$$
$$\Rightarrow \quad \rho_1 \tilde{Q}'(\rho_1) - \tilde{Q} = \frac{(\rho_1)^2(\kappa - 1)}{\kappa p}, (1 + p - 2\rho_1)$$

and thus we arrive at the desired results. Checking second order conditions in each variable yield that the function is convex in both arguments. As cross derivatives are 0 at the optimum, the critical point is a minimum by the second order derivative test. □

Proof of Lemma 3.6

Proof. **Step 1: The unconstrained optimum satisfies** (IC_i^l) **for** $p \leq 1/3$. As $U_i(l|h) = U_i(l|l)$ by Lemma 3.3 and with the help of Lemma B.3 stating that (IC_i^h) binds, we can rewrite (IC_i^l)

$$(\gamma_i(l) - \gamma_i(h))U_i(l|l) \geq (\gamma_i(l) - \gamma_i(h))U_i(h|h). \tag{B.20}$$

As $U_i(l|l) \geq U_i(l|h)$ by construction this holds if and only if $(\gamma_i(l) - \gamma_i(h)) > 0$.
Calculating the difference yields

$$\gamma_i(l) - \gamma_i(h) = \frac{p}{\rho_{-i}} \frac{\rho_i - p}{1 - p} \tag{B.21}$$

which is positive if and only if $\rho_i \geq p$.
Recall from Lemma 3.5 that the optimal unconstrained $\rho_2 = \frac{1-p}{2}$ which is larger p if and only if $p < 1/3$.

Step 2: Describe the equilibrium including (IC_i^l) **for** $p > 1/3$.

Step 2a: No solution with $\rho_1(l) > \rho_1(h)$. First, we show that we do not want to deviate to any $\rho_1(l) > \rho_1(h)$ for $p > 1/3$. To do so, consider (IC_2^l). By Lemma 3.3 the RHS remains at 0, and $U_2(l|h) > U_2(l|l)$. Thus, for (IC_2^l) to hold we would still need that $\gamma_2(l) \geq \gamma_2(h)$. However, then also $\tilde{\gamma_2}(l) - \tilde{\gamma_2}(h)$ needs to be positive. Plugging in and simplifying, we find that

$$\tilde{\gamma}_2(l) - \tilde{\gamma}_2(h) = \rho_1(h)\rho_2(l)(1 - p) - \rho_1(l)p^2(1 - \rho_2(l)) \tag{B.22}$$

which is decreasing in $\rho_1(l)$. Hence, no deviation to $\rho_1(l) > \rho_1(h)$ is profitable since whenever IC holds for this deviation, it also holds for $\rho_i(l) = \rho_i(h)$ which is preferred by Lemma 3.5.
Step 2b: The proposed solution is indeed an optimum. Next, we need to show that also no deviation to $\rho_1(l) < \rho_1(h)$ is optimal. For this we use a guess and verify approach to show that the proposed equilibrium with $\rho_2 = p$ is indeed an optimum.
To do this, this solution needs to satisfy the first order conditions of the Lagrangian at the proposed point. As we know from Step 2a we do not need to consider $\rho_1(l) >$

$\rho_1(h)$. Define $g(\boldsymbol{\rho}) \leq 0$ to be the incentive constraint, reformulated such that if $g \leq 0$, (IC_2^l) holds.[3] The Lagrangian is given by

$$\mathcal{L}(\lambda, \mu, \boldsymbol{\rho}) = Obj(\boldsymbol{\rho}) + \lambda g \boldsymbol{\rho}) + \mu(\rho_1(l) - \rho_1(h)). \tag{B.23}$$

Any solution to the constrained minimization problem $\boldsymbol{\rho}^*$ must be such that it solves the following problem

$$\min \mathcal{L}(\cdot) \tag{B.24}$$

and

$$\lambda, \mu \geq 0. \tag{B.25}$$

It turns out that the proposed solution is such a point and further $\mathcal{L}$ is strictly concave at this point, thus the problem is indeed locally minimized at $\boldsymbol{\rho}^*$.

Step 2c: Show that no other solution exists. It is not clear whether the problem is also globally minimized at this point, as both the objective as well as the constraint do not satisfy the usual assumption needed for global optimality, in particular they are not globally convex. However, fixing k we know the following two aspects:

(a) at $p = 1/3$ the solution is the same as the "unconstrained" optimum considered in Lemma B.3. For $p > 1/3$ the solution is worse than the unconstrained optimum,

(b) as all functions are continuous in p the functional value and thus the equilibrium value must be continuous in p.

This means that if another solution (strictly better than the candidate) exists for some $\hat{p} > 1/3$ then there also must exist some $\check{p} \in [1/3, \hat{p}]$ such that the equilibrium values $\hat{\boldsymbol{\rho}}$ of $\hat{p}$ as a function of p yield the same outcome as the proposed equilibrium. Further, as $\mathcal{L}$ is strictly convex at the proposed optimum, this alternative value $\hat{\boldsymbol{\rho}}$ must be bounded away in at least one of its variables.

Suppose the other optimal point is at some $\rho_i(k)$ not in the neighborhood of ρ_i^*. Then by continuity, the mean value theorem, and the strict convexity of $\mathcal{L}$ at the proposed point this point can only be optimal if the derivative of Obj w.r.t. $\rho_i(k)$ is 0 at some point on $(\rho_i(k), \rho_i^*)$.

As $\rho_1(l)$ has no extreme value on the interval $(\rho_2(l), \rho_1(h))$ by Step 1 in appendix B.3.5, $\rho_1(l)$ must be the same in both optima.

But then, if $\rho_1(l)$ is constant, $\rho_1(h)$ is increasing on $(a, 1)$. Then again $\rho_1(h) = 1$ cannot be optimal. Thus, no other minimum exists and our proposed minimum is the only and therefore global minimum. □

Proof of Lemma 3.7

Proof. Finally, introducing $\gamma(l, l) \leq 1$ to the problem it is straightforward to compute that the constraint has slack for any $p \geq 1/3$.

Also, by computing $\nu/(Q(P) - R(P))$ one can verify that it holds at ρ_1^*, ρ_2^* whenever

$$k \leq \frac{2 - 4p - 2p^2}{1 - 4p + 3p^2}.$$

[3] As $\rho_1 \geq p$ at the imposed constrained optimum, we do not worry about (IC_1^l) which always has slack.

Further, if the constraint $\gamma(l,l) \leq 1$ binds, we can use Lemma B.4 to see that $\gamma(l,h), \gamma(h,l) \leq 1$ if and only if $\rho_i(l) \geq p$.
We know that at the unconstrained optimum with $\rho_1(l) = \rho_1(h)$ and thus, we have a boundary solution in those variables for a given $\rho_2(l)$. However, the solution with respect to $\rho_2(l)$ is such that $Obj'(\rho_2(l)) = 0$.
In addition we know by strict concavity that in fact the regime change happening at $\rho_1(l) = \rho_1(h)$ (from high-cost types having a beneficial deviation payoff to low cost types having one), must be such that around the unconstrained optimum we would not change the equation $\rho_1(l) = \rho_1(h)$ as this would either provide us with a free lunch lowering $\rho_1(h)$ to put slack on $\gamma(l,l) \leq 1$. Then, as we change the regime to $\rho_1(l) > \rho_1(h)$ it must be that $Obj'(\rho_1(l)) > 0$ as we started at the optimum. Thus, we could lower $\rho_1(l)$ at no cost on the constraint to $\rho_1(l) = \rho_1(h)$ as the constraint can be rewritten as

$$\nu/(Q(P) - R(P)) - 1 = Obj - R \leq 0,$$

and $R|_{\rho_1(l)=\rho_1(h)} = p^2/\rho_1(l)\rho_2(l)$.
As $\rho_1(l) = \rho_1(h)$ remains to hold the problem

$$\min_{\rho_1,\rho_2} Obj$$

s.t. $\rho_2 \geq p$ and $\gamma(l,l) \leq 1$ is well-behaved such that we get the desired solution of the lemma.
Finally, plugging the solution for every regime into the Border constraints (EPI) and (IF) shows that they hold at the optimum. □

B.3.6 Proof of Proposition 3.2

Proof. For the first result, observe that for $p < 1/3$, the solution at which $\rho_2 = p$ and $\rho_1 = 2p + 1/(\kappa - 1)$ is always feasible and in line with $\gamma(l,l) \leq 1$ and (IC_i^l). The corresponding probability of contest is given as

$$Pr(\Gamma, \boldsymbol{\rho}^*) = \frac{(\kappa+1)p}{1 + 2(\kappa-1)p}, \tag{B.26}$$

which is increasing in p and κ and becomes $1/2$ for $p = 1/3$ and $\kappa \to \infty$.
Second, the optimal probability of a contest for $p > 1/3$ is

$$4p\frac{\kappa p - (1-p)(\kappa-2)}{(\kappa-1)(7p^2 - 2p - 1) + 4p}, \tag{B.27}$$

which is falling in p for $p > 1/3$. Thus, it suffices to look at the probability at $p = 1/3$. But at this point it becomes

$$\frac{\kappa - 4}{2\kappa - 5}, \tag{B.28}$$

which again is bounded by $1/2$.
The inverse u-shape follows from $Pr(\Gamma, \boldsymbol{\rho}^*)$ being concave on all intervals and that the derivative is smooth pasting at p', p'', p'''.

Finally, monotonicity (and concavity) in κ follows from monotonicity and concavity in κ for all regions as well as smooth pasting at the transition of the regions. □

B.3.7 Proof of Proposition 3.3

Proof. The results for the probability of being send to contest follow immediately from the ex-ante symmetry and the equilibrium beliefs specified in Theorem 3.1 and 3.7.

The result on the expected share follows from Theorem 3.1 and Lemma 3.1. The low-cost types expected utility from contest is weakly below her outside option V. In order to fulfill the participation constraint in expectations, the player needs to be compensated by a higher share if mediation fails. As player $1l$ has a higher probability to enter the contest, she also needs to receive a higher share than player $2l$. A weakly higher share for any l-type compared to the same player's corresponding h-type follows from h-types binding incentive compatibility. Finally, as player $1h$ gains a positive expected utility in case of the contest her expected share can be pushed down the most completing the proof. □

B.3.8 Proof of Proposition 3.4

Proof. The expected legal expenditure of player ik is by the uniform equilibrium scoring functions given by

$$E[LE_i^k] = \sum_{r\in\{b,m,t\}} Prob(s_i^k \in r)\frac{\underline{r}+\overline{r}}{2}$$

where b, m and t are the scoring ranges used in Figure 3.1 and the proof of Lemma 3.1. Further, $\underline{r}$ denotes the upper bound of range r and $\underline{r}$ denotes the lower bound of range r.

The expected scoring function of player 1 entirely depends on ρ_2, that is

$$\begin{aligned}\rho_1 E[LE_1^l] + (1-\rho_1)E[LE_1^h] &= \rho_1\frac{\rho_1(2-\rho_1)+(\rho_2)^2(\kappa-1)}{2\rho_1 c_l\kappa} + (1-\rho_1)\frac{(1-\rho_1)}{2c_l\kappa}\\ &= \frac{1+\rho_2(\kappa-1)}{c_l\kappa}.\end{aligned}$$

Thus, the equilibrium expected contest score of player 1 is the same as in a contest without mediation whenever $\rho_2 = p$.

The expected score of player 2 is computed in a similar manner but depends on both ρ_1 and ρ_2. It is given by:

$$\frac{1}{2c_l\kappa}\left(\frac{(\kappa-1)}{\kappa}\Big(\rho_1(\rho_1-2)+(\rho_2)^2(\kappa-1)+2\rho_2\Big)+1\right).$$

The derivative of this function w.r.t. to ρ_1

$$\frac{\kappa-1}{\kappa^2}(\rho_1-1) < 0.$$

As $\rho_1 > p$ by Lemma 3.7 and $\rho_2 = p$ for $p \notin (p', 1/3)$, it follows that total legal expenditures post-mediation are indeed smaller than under the prior belief p. □

B.3.9 Proof of Proposition 3.5

Proof. As participation is optimal by lemma B.1 and the optimal mechanism is unique, no bargaining protocol can achieve a better result than Theorem 3.1. By convexity of contest utilities in beliefs, no Bayes plausible signal structure over the prior can make the receiver worse-off than the prior. Thus, the participation constraint of the mechanism holds in the bargaining game as well.

To show that take-it-or-leave-it bargaining performs worse in environments that satisfy 1 we show that the low-cost type of Sender has always an incentive to deviate to some offer $0 < \epsilon < 1 - V(l)$ that yields a utility higher than $V(l)$ which is her on-path utility. We do so by considering the possible response of Receiver to such an offer given any off-path β_S describing the probability assessment of Receiver on Sender in the contest game.

Any Receiver type accepts. As $\epsilon < 1 - V(l)$, Sender earns a utility larger $V(l)$.

Any Receiver type rejects. The high-type only rejects an offer of ϵ if she expects a utility $U_R(h|\beta_S) > \epsilon$, given her off-path belief β_S. By Lemma 3.1 $U_R(h|\beta_S, \beta_R) > 0$ only if $\beta_S < \beta_R$. Since any Receiver type rejects the offer, the belief on the receiver is the same as the prior, that is $\beta_R = p$. But $\beta_S < p$ implies via lemma 3.1 that $U_S(l|\beta_S, \beta_R) > V(l)$.

h-type Receiver rejects and l-type Receiver accepts. This case doesn't exist, as any offer that the h-type rejects is also rejected by the l-type as PBE requires type-independent beliefs after the deviation (Fudenberg and Tirole, 1988) and l-types have lower cost of evidence provision.

l-type Receiver rejects and h-type Receiver accepts. h-types only accept if $\epsilon \geq U_R(h|\beta_S, \beta_R)$ that is

$$\epsilon \geq (\beta_R - \beta_S)\frac{\kappa - 1}{\kappa}.$$

If Receiver h, type l rejects, then Sender, type l gains $(1-p)(1-\epsilon)$ which is larger $V(l) = (1-p)(\kappa - 1)/\kappa$ as ϵ goes to 0. Thus Receiver, type h must be indifferent. Rewriting the above equation yields

$$\beta_R = \frac{\epsilon\kappa}{\kappa - 1} + \beta_S.$$

In order to induce a belief of β_R, Receiver, type h must choose to reject the offer with probability

$$\gamma_{R,h} = \frac{p}{1-p}\frac{1-\beta_R}{\beta_R},$$

which follows analogously to Lemma B.4.

Plugging this into Sender l-types yields:

$$(1-p)(1-\gamma_{R,h})(1-\epsilon) + (p + (1-p)\gamma_{R,h})(1-\beta_S)\frac{\kappa-1}{\kappa} =$$
$$(1-p)(1-\epsilon) + \frac{p}{\beta_R}\left((1-\beta_S)\frac{\kappa-1}{\kappa} - (1-\beta_R)(1-\epsilon)\right).$$

Taking into account that β_R is a function of β_S this expression is continuous and monotone in β_S. β_S is naturally bounded by 1 and β_R. As we are looking for the lowest utility, we can assign for any $\epsilon > 0$ it suffices to consider an upper and a lower bound. For ϵ close to 0 however, both $\beta_S = \beta_R$ as well as $\beta_S = 1$ yield a utility larger $(1-p)(\kappa-1)/\kappa$. Thus, Sender, type l always has an incentive to deviate to some ϵ irrespective of the out-of-equilibrium beliefs of Receiver resulting in an inferior solution which is actually strict as long as the case is not trivial by the uniqueness of the proposed mechanism as shown in the proof of Theorem 3.1. □

B.3.10 Proof of Proposition 3.6

The proof relies on three features of the model which can be exploited to guarantee a weaker participation constraint:

- the mediator can ex-ante commit to probabilistic private messages she sends to parties following any given message profile (but before the acceptance decision),
- the mediator can ex-ante commit to an additional probabilistic private message she sends to parties following any message and acceptance profile (that is after the acceptance decision),
- all type profiles lead to on-path to litigation with positive probability.

Proof. For PAIR we need that the expected share given one's own type, that is $x_i(l)$ is larger than the expected utility of a contest that occurs upon rejection of this share. Suppose without loss of generality that an offer of 0 is rejected by all parties and is used by the mediator to trigger litigation.

Two aspects facilitate the analysis: First, the mediator can choose a signal $\sigma(\mathbf{m}, \mathbf{d})$ that depends on the received messages $\mathbf{m}$ as well as on the acceptance decision $\mathbf{d}$ of both players. That is, the mediator has the possibility to define a post-mediation protocol, too.

Recall from Theorem 3.1 that any type profile leads to litigation with positive probability. At the same time rejection by one party is enough to trigger litigation. Thus, as we allow for private communication, the mediator is free to choose one of the two messages sent to one party if she triggers rejection by the other party. The mediator can therefore randomize not only between who takes the role of player 1, that is which $\mathcal{X}_i$ to use, but also between whom of the two player's receives the "trigger message" 0. For the non-triggering player the mediator can in fact randomizes between all messages the player could receive on-path when the conflict is settled. This way the player does not know whether she is treated as player 1 or player 2 in the mediation protocol at the time of making her decision as to whether to accept or reject the offer. She does in fact not even know whether rejecting the offer makes any difference at all (as the opponent might have received an offer of 0 anyways). By Proposition 3.3, the mediator can choose $\mathcal{X}_i$ such that for any offer $x_i(k)$ there exists an on-path continuation game in which the player is worse off than $x_i(k)$. Hence, it is possible for the mediator to choose a signal σ_i conditional on deviation that signalling the deviator is in this on-path subgame deterring deviation altogether. □

B.3.11 Proof of Proposition 3.7

Proof. Whenever $\gamma_i(k) \neq 1$ the proof is the same as that of Proposition 3.6.

The situation is however different if either of the players is sent to court with probability $\gamma_i(k) = 1$. According to Theorem 3.1 and Lemma 3.7, $\gamma_i(h) < 1$. In addition at most one of the l-types has $\gamma_i(l) = 1$ on path.
This way the player knows that in one of the two mediation protocols she is always going to litigate anyways. Thus if $x_1(l,h) \neq x_1(l,l)$, player 1 might have a strong incentive to deviate as she knows whom she is facing in case her decision is relevant at all. In all other cases she is going to litigate anyways and receives $V(l)$ as litigation payoff by Theorem 3.1 together with Lemma 3.1. Thus, it might be optimal for her to reject anything but $x_1(l,l)$.
Suppose instead the mediator announces a mediation protocol $\mathcal{X}_\lambda^\epsilon$ in which reporting two l-types follows mediation breakdown with full information disclosure with probability ϵ and a protocol as that derived in Section 3.4 otherwise. As $\epsilon \to 0$, the result gets arbitrarily close to that of Theorem 3.1. However, the mediator can signal any l-type deviator that in fact the low-cost vs. low-cost litigation game is played, causing the l-type to also except ex-post shares. □

B.3.12 Proof of Lemma B.6

Proof. If condition (M) is violated, the equilibrium is no-longer monotonic but instead overlapping strategies might be possible. The reason for this is that if, e.g. $p_1(l|l)\kappa < p_1(l|h)$ the likelihood of meeting a low-cost type when being a high-cost type is too high compared to being a low-cost type, such that the high-cost type has a strong incentive to *overscore* the low-cost type. Further, by the consistency condition equation (C) whenever the high-cost type faces a low-cost type, she faces indeed a low-cost type that *thinks* she herself is facing a high-cost type with very high probability. This provides an incentive for the h-type to compete more aggressive and for the l-type to compete softer than under condition (M). The equilibrium scores in the non-monotonic equilibrium are as depicted in figure B.1. Player $1l$ and player $1h$ overlap on the middle interval but are otherwise "close to monotonic". While the high-cost type of player 2 has a support covering the whole scoring interval, player $2l$ only competes in the middle interval. In addition player $2h$ also has a mass point at 0. Solving for the optimal mechanism, it turns out that

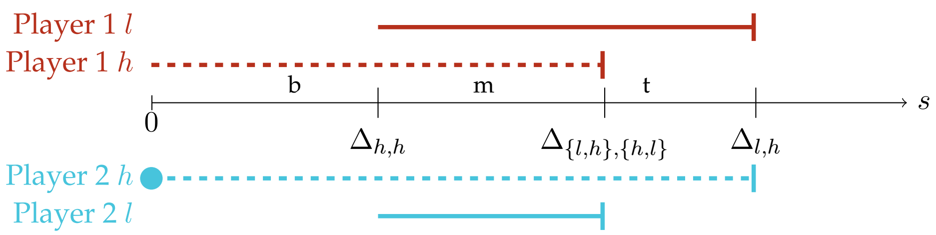

FIGURE B.1: *Strategy support of player 1 and 2 if monotonicity fails.*

there is still no interior solution in $p_1(l|l)$. The mediator would set $p_1(l|l)$ equal to any discontinuity point or at the respective borders. That is either $p_1(l|l) = 0$ or $p_1(l|l) = \max\{p_2(l|l), p_1(l|h)/\kappa\}$. If $p_1(l|l) = p_2(l|l) = \rho(l)$ under non-monotonicity, the first order condition of the mediator's problem is monotone in $\rho(l)$ and thus, we would need $\rho(l) = 0$ which is never optimal. If $p_1(l|l) = p_1(l|h)/\kappa$ utilities converge to their monotone counterparts and thus, the solution is no different than that for monotonicity. Finally, $p_1(l|l) = 0$ is never optimal as the objective is always decreasing at this point. □

Bibliography

Anderson, Heather and Ron Pi (2004). *Evaluation of the Early Mediation Pilot Programs*. Tech. rep. Judical Council of California.
Aoyagi, Masaki (2003). "Bid rotation and collusion in repeated auctions". In: *Journal of Economic Theory* 112.1, pp. 79–105.
— (2007). "Efficient collusion in repeated auctions with communication". In: *Journal of Economic Theory* 134.1, pp. 61–92.
Asker, John (2010). "A study of the internal organization of a bidding cartel". In: *The American Economic Review* 100.3, pp. 724–762.
Baye, Michael R, Dan Kovenock, and Casper G Vries (1996). "The all-pay auction with complete information". In: *Economic Theory* 8.2, pp. 291–305.
— (2005). "Comparative Analysis of Litigation Systems: An Auction-Theoretic Approach". In: *The Economic Journal* 115.505, pp. 583–601.
Bebchuk, Lucian Arye (1984). "Litigation and settlement under imperfect information". In: *The RAND Journal of Economics*, pp. 404–415.
Bergemann, Dirk, Benjamin A Brooks, and Stephen Morris (2013). "Extremal Information Structures in the First Price Auction". In:
Bester, Helmut and Karl Wärneryd (2006). "Conflict and the Social Contract". In: *The Scandinavian Journal of Economics* 108.2, pp. 231–249.
Bichler, Martin et al. (2006). "Industrial procurement auctions". In: *Combinatorial Auction*.
Border, Kim C (2007). "Reduced form auctions revisited". In: *Economic Theory* 31.1, pp. 167–181.
Börgers, Tilman and Peter Norman (2009). "A note on budget balance under interim participation constraints: the case of independent types". In: *Economic Theory* 39.3, pp. 477–489.
Brown, Jennifer Gerarda and Ian Ayres (1994). "Economic rationales for mediation". In: *Virginia Law Review*, pp. 323–402.
Brown, Scott, Christine Cervenak, and David Fairman (1998). *Alternative Dispute Resolution Practitioners Guide*. Tech. rep. Conflict Management Group.
Cave, Jonathan and Stephen W Salant (1987). "Cartels that vote: Agricultural marketing boards and induced voting behavior". In: *Public Regulation*. Ed. by E. Bailey. MIT Press.
Celik, Gorkem and Michael Peters (2011). "Equilibrium rejection of a mechanism". In: *Games and Economic Behavior* 73.2, pp. 375–387.
Che, Yeon-Koo and Ian Gale (2000). "Difference-form contests and the robustness of all-pay auctions". In: *Games and Economic Behavior* 30.1, pp. 22–43.
Che, Yeon-Koo and Jinwoo Kim (2006). "Robustly Collusion-Proof Implementation". In: *Econometrica* 74.4, pp. 1063–1107.
Cho, In-Koo and David M Kreps (1987). "Signaling games and stable equilibria". In: *The Quarterly Journal of Economics*, pp. 179–221.

Cramton, Peter C and Thomas R Palfrey (1995). "Ratifiable mechanisms: learning from disagreement". In: *Games and Economic Behavior* 10.2, pp. 255–283.
Dequiedt, Vianney (2006). *Ratification and veto constraints in mechanism design*. Tech. rep.
Doornik, Katherine (2014). "A rationale for mediation and its optimal use". In: *International Review of Law and Economics* 38, pp. 1 –10. ISSN: 0144-8188. DOI: http://dx.doi.org/10.1016/j.irle.2013.11.005. URL: http://www.sciencedirect.com/science/article/pii/S0144818813000859.
Eső, Péter and James Schummer (2004). "Bribing and signaling in second price auctions". In: *Games and Economic Behavior* 47.2, pp. 299–324.
Ewerhart, Christian (2015). "Rent-seeking games and the all-pay auction". In: *mimeo* 186.
Fey, Mark and Kristopher W. Ramsay (2010). "When Is Shuttle Diplomacy Worth the Commute? Information Sharing through Mediation". In: *World Politics* 62 (04), pp. 529–560. ISSN: 1086-3338. DOI: 10.1017/S0043887110000183. URL: http://journals.cambridge.org/article_S0043887110000183.
Francetich, Alejandro and Pete Troyan (2012). "Collusive Agreements in Auctions: Design and Execution by an Informed Principal". In:
Fudenberg, Drew and Jean Tirole (1988). "Perfect Bayesian and Sequential Equilibria - A clarifying Note". In: *mimeo*.
Garratt, Rodney J, Thomas Tröger, and Charles Z Zheng (2009). "Collusion via resale". In: *Econometrica* 77.4, pp. 1095–1136.
Genn, Hazel (1998). *The Central London County Court Pilot Mediation Scheme*. Tech. rep. Department for Constitutional Affairs.
Graham, Daniel A and Robert C Marshall (1987). "Collusive bidder behavior at single-object second-price and English auctions". In: *The Journal of Political Economy*, pp. 1217–1239.
Hörner, Johannes, Massimo Morelli, and Francesco Squintani (2015). "Mediation and Peace". In: *The Review of Economic Studies* 82.4, pp. 1483–1501. DOI: 10.1093/restud/rdv022. eprint: http://restud.oxfordjournals.org/content/82/4/1483.full.pdf+html. URL: http://restud.oxfordjournals.org/content/82/4/1483.abstract.
Jullien, Bruno (2000). "Participation constraints in adverse selection models". In: *Journal of Economic Theory* 93.1, pp. 1–47.
Katz, Avery (1988). "Judicial decisionmaking and litigation expenditure". In: *International Review of Law and Economics* 8.2, pp. 127–143.
Koessler, Frédéric and Vasiliki Skreta (2014). "Sales Talk". In: *Available at SSRN 2465174*.
Krishna, Vijay (2002). *Auction theory*. Academic press.
Laffont, Jean-Jacques and David Martimort (1997). "Collusion under asymmetric information". In: *Econometrica: Journal of the Econometric Society*, pp. 875–911.
— (2000). "Mechanism design with collusion and correlation". In: *Econometrica* 68.2, pp. 309–342.
Ledyard, John O and Thomas R Palfrey (2007). "A general characterization of interim efficient mechanisms for independent linear environments". In: *Journal of Economic Theory* 133.1, pp. 441–466.
Llorente-Saguer, Aniol and Ro'i Zultan (2014). "Auction Mechanisms and Bidder Collusion: Bribes, Signals and Selection". In: *MPI Collective Goods Preprint* 2014/18.

Lopomo, Giuseppe, Robert C Marshall, and Leslie M Marx (2005). “Inefficiency of collusion at english auctions”. In: *Contributions in Theoretical Economics* 5.1.

Lopomo, Giuseppe, Leslie M Marx, and Peng Sun (2011). “Bidder collusion at first-price auctions”. In: *Review of Economic Design* 15.3, pp. 177–211.

Mailath, George J and Peter Zemsky (1991). “Collusion in second price auctions with heterogeneous bidders”. In: *Games and Economic Behavior* 3.4, pp. 467–486.

Marshall, Robert C and Leslie M Marx (2007). “Bidder collusion”. In: *Journal of Economic Theory* 133.1, pp. 374–402.

— (2009). “The vulnerability of auctions to bidder collusion”. In: *The Quarterly Journal of Economics* 124.2, pp. 883–910.

Marshall, Robert C, Leslie M Marx, and Michael J Meurer (2014). “The economics of bidder collusion”. In: *Game Theory and Business Applications*. Springer, pp. 367–397.

Marshall, Robert C et al. (1994). “Numerical analysis of asymmetric first price auctions”. In: *Games and Economic Behavior* 7.2, pp. 193–220.

Maskin, Eric and Jean Tirole (1992). “The principal-agent relationship with an informed principal, II: Common values”. In: *Econometrica: Journal of the Econometric Society*, pp. 1–42.

McAfee, R Preston and John McMillan (1992). “Bidding rings”. In: *The American Economic Review*, pp. 579–599.

Meirowitz, Adam et al. (2015). “Dispute Resolution Institutions and Strategic Militarization”. In: *mimeo*.

Milgrom, Paul and Ilya Segal (2002). “Envelope theorems for arbitrary choice sets”. In: *Econometrica* 70.2, pp. 583–601.

Mitusch, Kay and Roland Strausz (2005). “Mediation in Situations of Conflict and Limited Commitment”. In: *Journal of Law, Economics, and Organization* 21.2, pp. 467–500. DOI: `10.1093/jleo/ewi018`.

Myerson, Roger B (1983). “Mechanism design by an informed principal”. In: *Econometrica: Journal of the Econometric Society*, pp. 1767–1797.

Myerson, Roger B and Mark A Satterthwaite (1983). “Efficient mechanisms for bilateral trading”. In: *Journal of economic theory* 29.2, pp. 265–281.

Mylovanov, Tymofiy and Thomas Tröger (2015). “The sender-payoff approach to signaling and the informed-principal problem”. In: *mimeo, https://thomasetroger.wordpress.com/eine-seite/research/*.

Pavlov, Gregory (2013). “Correlated equilibria and communication equilibria in all-pay auctions”. In: *mimeo*.

Posner, Richard A (1973). “An economic approach to legal procedure and judicial administration”. In: *The Journal of Legal Studies*, pp. 399–458.

— (1996). *The Federal Courts: Challenge and Reform*. Harvard University Press.

Prescott, J J, Kathryn E Spier, and Albert Yoon (2014). “Trial and Settlement: A Study of High-Low Agreements”. In: *Journal of Law and Economics* 57, pp. 699–821.

Quesada, Lucia (2004). “Collusion as an informed principal problem”. In: *University of Wisconsin*.

Rachmilevitch, Shiran (2013). “Bribing in first-price auctions”. In: *Games and Economic Behavior* 77.1, pp. 214–228.

Robinson, Marc S (1985). “Collusion and the Choice of Auction”. In: *The RAND Journal of Economics*, pp. 141–145.

Schweizer, Urs (1989). “Litigation and settlement under two-sided incomplete information”. In: *The Review of Economic Studies* 56.2, pp. 163–177.

Shavell, Steven (1995). “Alternative dispute resolution: an economic analysis”. In: *The Journal of Legal Studies*, pp. 1–28.

Siegel, Ron (2009). “All-Pay Contests”. In: *Econometrica* 77.1, pp. 71–92.

— (2014). “Asymmetric all-pay auctions with interdependent valuations”. In: *Journal of Economic Theory* 153.0, pp. 684 –702. ISSN: 0022-0531. DOI: http://dx.doi.org/10.1016/j.jet.2014.03.003. URL: http://www.sciencedirect.com/science/article/pii/S0022053114000295.

Spier, Kathryn E. (1994). “Pretrial bargaining and the design of fee-shifting rules”. In: *The RAND Journal of Economics*, pp. 197–214.

— (2007). “Litigation”. In: *Handbook of law and economics* 1, pp. 259–342.

Spier, Kathryn E and David Rosenberg (2011). “A Contest Model of Separate Action versus Class Action Litigation”. In: *Journal of Legal Analysis*, pp. 11–28.

Szech, Nora (2011). “Asymmetric all-pay auctions with two types”. In: *mimeo*.

Tan, Guofu and Okan Yilankaya (2007). “Ratifiability of efficient collusive mechanisms in second-price auctions with participation costs”. In: *Games and Economic Behavior* 59.2, pp. 383–396.

Tullock, Gordon (1980). “Efficient rent seeking”. In: *Towards a Theory of the Rent-Seeking Society*. Ed. by James Buchanan, Gordon Tullock, and Roger Tollison. College Station, TX: Texas A and M University Press, pp. 97–112.

Ungern-Sternberg, Thomas von (1988). “Cartel stability in sealed bid second price auctions”. In: *The Journal of Industrial Economics*, pp. 351–358.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig angefertigt und die benutzen Hilfsmittel vollständig und deutlich angegeben habe.

Mannheim, 31. Mai 2016

Benjamin Balzer