Belief Management and Optimal Arbitration[*]

Benjamin Balzer[†] Johannes Schneider[‡]

September 20, 2019

Abstract

We consider a general model of arbitration. The outcome following failed settlement is endogenous and depends on the arbitration mechanism. We show that the arbitration problem reduces to an information-design problem of finding the optimal information structure upon failure to settle. The result holds for a wide range of arbitration objectives. Our environment is characterized by five axioms: ordered types, desirable settlement, unilateral veto rights, exogenous rules of conflict, and budget balance. We nest several models from the literature. The information-design problem characterizes optimal arbitration as a function of the rules of the game that follows failed arbitration attempts.

[*] We thank Eddie Deckel, Piotr Dworczak, Willie Fuchs, Johannes Hörner, Antoine Loeper, Benny Moldovanu, Volker Nocke, Alessandro Pavan, Thomas Tröger, and Christoph Wolf for helpful comments at various stages of the paper. Some parts of this paper have earlier been circulated as part of another paper with title "Managing a Conflict." Johannes Schneider gratefully acknowledges support from the Ministerio Economia y Competitividad, Maria de Maeztu grant (MDM 2014-0431), and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444).

[†] University of Technology Sydney, benjamin.balzer@uts.edu.au

[‡] Carlos III de Madrid, Department of Economics, jschneid@eco.uc3m.es

Contents

1 Introduction

Resolving conflicts through an open fight often implies high costs. Thus, attempting to resolve conflicts before they escalate to a fight is common. One of the most powerful resolution attempts is third-party arbitration. Once parties agree on arbitration, an arbitrator controls the outcome of arbitration. Still arbitrators seldom guarantee settlement and often operate *in the shadow of the fight*.

In this paper, we propose a general framework that allows for an arbitrary game in the fighting stage to answer the question: *How should we design arbitration?* In particular, we address how the details of the fighting stage affect optimal arbitration.

We develop an information-design problem to determine optimal information revelation within arbitration. We show that the solution to that problem is necessary and sufficient to determine optimal arbitration. The problem admits an intuitive formulation for a range of objectives and is directly defined on the properties of the game in the fighting stage. It provides a direct link between these properties and optimal arbitration. The mapping from optimal information revelation to the optimal mechanism is the solution to a system of linear equations.

Our results provide an intuition why the features of the fighting stage are of first-order importance to design optimal arbitration. Moreover, they determine the two major channels: It is key to understand the effect of information revelation on *joint surplus* and on *the payoff relevance of a player's private type*. If either is sensitive to information revelation, the optimal mechanism is shaped by the need to manage players' updating behavior. Moreover, if *both* effects are sensitive to information revelation, the designer faces a trade-off to balance them.

Our result identifies how the arbitrator's objective influences optimal arbitration. We show that even if the designer only cares about settlement solutions, she has an incentive to increase surplus. In turn, a designer that only cares about surplus has an incentive to reduce surplus sometimes to screen more effectively.

We show that the optimal mechanism is robust to the arbitrator's objective if either joint surplus or the impact of players' private types in the fighting stage is constant in the information structure. If both are sensitive to information, optimal arbitration is sensitive to the chosen objective. Indirect incentives to the arbitrator may lead to distortions.

We set up a general model of arbitration (Section 2) that covers a variety of settings. We nest classic reduced-form cases from the literature, but also allow for settings in which information revealed in arbitration has strategic relevance in the fighting stage. The latter cases complicate the problem. Players are aware of the strategic value of information. They have an increased privacy concern and an incentive to manipulate the information structure. Optimal arbitration takes both effects into account. We highlight that the information revelation in arbitration is of first-order importance.

In the first part of the paper (Section 3) we argue that our model captures *canonical arbitration problems*. We formulate five axioms: (i) private information about strength in the fight that can be ordered; (ii) efficient settlement; (iii) unilateral veto rights; (iv) exogenous rules of the fighting stage; and (v) arbitration without a structural deficit.

These axioms are the common features of arbitration problems both in reality and in the existing literature. The appropriate objective function of an arbitrator depends heavily on the environment.[1] We assume that the arbitrator has (weakly) monotone preferences both in cases settled and in joint surplus. The assumption nests maximizing joint surplus and maximizing cases settled as special cases.

We follow the literature in setting up the problem as a classic mechanism-design problem. If the type distribution matters once the conflict escalates, the design of the mechanism has an effect on the outcome of the (potential) fight. If that effect is non-trivial, it turns out that classic mechanism-design techniques are often impractical (i) to disentangle the mechanism-design part from the information-design part, (ii) to isolate the arbitrator's trade-off, and (iii) to compute the solution.

Most approaches in the literature make specific assumptions on the effect of information revelation. In particular they assume either *no or only non-strategic* effects of information revelation on players' behavior in the fighting stage.[2] In a related paper Balzer and Schneider (2019) we show by means of a particular application that conclusions are sensitive to that assumption.

In the second part of the paper (Section 4) we reduce the problem to a simpler, yet equivalent one: the belief-management problem (Theorem 1). It is an information-design problem directly defined on the rules of the fighting stage. Its solution implies the solution to the arbitration problem. Optimal belief management determines the information structure in the fighting stage.

[1] In military disputes for example, the appropriate objective of the UN may be to minimize the likelihood of war, while OPEC's cartel officer aiming to settle a dispute within the organization may have the joint surplus of its members as the objective.

[2] Examples include Spier (1994), Bester and Wärneryd (2006), Mylovanov and Zapechelnyuk (2013), Hörner, Morelli, and Squintani (2015), and Meirowitz et al. (2017) among others. Recent exceptions are Zheng (2018), Zheng and Kamranzadeh (2018), and Balzer and Schneider (2019).

The construction applies whenever full settlement is not possible. The one-to-one mapping from information structures to candidate mechanisms is characterized by a set of linear equations and thus straightforward to compute.

Optimal belief management is necessary and sufficient for optimal arbitration. Belief management allows us (i) to disentangle the mechanism-design part from the information-design part, (ii) to isolate and interpret the arbitrator's trade-off, (iii) and to reduce the computational complexity.

To further characterize optimal arbitration we restrict the designer's objective to a weighted average between maximizing surplus and maximizing settlement. In this class of objectives, belief-management problems have a clear economic interpretation. The arbitrator seeks to implement fights with a high level of fundamental discrimination and a low level of inefficiency. We develop a formulation of the information-design problem that is easy to interpret and provides the main economic intuition.

In the *last part of the paper* (Section 5) we illustrate the benefits of our approach by means of specific examples. We provide a reinterpretation of the existing literature and show how to relate the properties of the game in the fighting stage to the arbitration problem.

Related Literature. The literature on arbitration is extensive and results differ significantly between models. We aim at providing a model that nests any setting in which private information is the main obstacle to settlement (Brown and Ayres, 1994).

We relate to Mylovanov and Zapechelnyuk (2013) who address mandatory arbitration from a design perspective.[3] With slight modifications their environment is nested by ours (see our discussion around Proposition 4 and 5). The optimal mechanism provides full settlement which is unsurprising in light of our Corollary 1. In addition they compare the implementation of specific mechanisms and find that if the arbitrator cannot control *when* to escalate, the mechanism is inferior. Our findings provide a rationale for that result. We show (for the general case) that it is indeed of first-order relevance to control the players' information sets and thus the escalation choice.

Another strand of the arbitration literature considers cases with binding participation constraints. Both the law and economics literature (Bebchuk, 1984; Schweizer, 1989; Spier, 1994, and the literature following), and the literature on international conflicts (Bester and Wärneryd, 2006; Fey and Ramsay, 2011;

[3] See also Armstrong and Hurley (2002) and Olszewski (2011) and references therein.

Jackson and Morelli, 2011; Hörner, Morelli, and Squintani, 2015, and the literature following) consider arbitration mechanisms. A common feature is that the information obtained during arbitration has no effect on players' decisions once settlement fails. We fully nest these models.

Models on arbitration where information revelation and continuation play interact are rare. Our own work on alternative dispute resolution (Balzer and Schneider, 2019) considers an all-pay auction as the alternative to arbitration. In that paper we apply the methods developed here (among others) to fully characterize optimal alternative dispute resolution in the legal system.[4]

Similar to us, models on common agency by Calzolari and Pavan (2006a,b) and Pavan and Calzolari (2009) emphasize that the design choices within a mechanism affect action choices outside that mechanism. The common theme in the large literature on resale (e.g. Gupta and Lebrun, 1999; Zheng, 2002; Goeree, 2003; Carroll and Segal, 2018) is that information revelation affects the outcome of an auction. The literature on aftermarkets (Lauermann and Virág, 2012; Atakan and Ekmekci, 2014; Zhang, 2014; Dworczak, 2017) follows Calzolari and Pavan (2006a) by looking at the interaction between the design of the mechanism and action choices in the aftermarket. However, almost all of that literature makes detailed assumptions on either the mechanism or the disclosure rule. An exception is Dworczak (2017), the paper closest to ours in that literature.

Conceptually, there are two major difference between arbitration and aftermarkets: First, in arbitration the fighting stage serves as the main screening instrument, while the potential to resell an object is often an obstacle to screening. Second, in arbitration the set of players is identical inside and outside the mechanism. As a consequence, in our model all players learn from arbitration about their opponents' type and interpret the information in light of their own history of play. In contrast, in the aftermarket literature players can only learn about those opponents that participated in the mechanism. This restricts the potential information structure under which the aftermarket is played.

On an abstract level the arbitration problem is a mechanism-design problem with adverse selection and moral hazard a la Myerson (1982). It involves information externality in the sense of Jehiel and Moldovanu (2001). Contrary to their paper the externality also affects the players' behavior after the mechanism in our setting.

[4] Zheng and Kamranzadeh (2018) consider a model identical to the baseline in Balzer and Schneider (2019). They restrict attention to take-it-or-leave-it settlement offers. Zheng (2018) assumes an all-pay auction as the fighting game and a determines when full settlement can be guaranteed.

Our belief-management representation transforms the mechanism-design problem into an information-design problem a la Bergemann and Morris (2016). Recent techniques proposed by Mathevet, Perego, and Taneva (2017), Dworczak and Martini (2018), Galperti and Perego (2018), Kolotilin (2018), and Kolotilin and Zapechelnyuk (2019) can be applied to address the belief-management problem.

We relate to these models in another dimension too. Like us, they determine the price of implementing an information structure. However, and different to us, they treat the prior distribution as the designer's initial, exogenous endowment. In our model, the designer endogenously "produces" that endowment. The relation between the mechanism-design part and the information-design part determines the *cost* of producing the "prior" of the fighting stage.[5] The arbitrator has a hybrid task: acquire information and disseminate it.

We relate to recent work by Georgiadis and Szentes (2018) that considers optimal monitoring rules. While different from our model in many dimensions, they, too, obtain an information-design formulation that adds tractability by separating the contract-design part from the information-design part.

2 Model

In this section we provide a general model of arbitration. Our model allows for a large class of escalation games and nests several arbitration models from the literature (or their discretized version) as special cases.[6] We revisit the examples from the literature in Section 5.1.

2.1 Model Primitives

Players and Types. There are two ex-ante symmetric, risk-neutral players, A and B. Each player i has a privately known *type* $\theta_i \in \Theta \equiv \{1, 2, ..., K\}$. Types are independently distributed according to a distribution $p : \Theta \to [0, 1]$ with $\sum\limits_{\theta=1}^{K} p(\theta) = 1$. The symmetry assumption is for notational convenience only.

Final Outcomes. The final outcome depends on which of the following *disjoint events* realizes: veto, $\mathcal{V}$, escalation, $\mathcal{E}$, or settlement, $\mathcal{Z}$. We define outcomes separately.

[5] Dworczak (2017) determines a similar object (the "no-communication posterior") for the case of single-agent mechanisms with aftermarkets.

[6] Examples are Spier (1994), Bester and Wärneryd (2006), Mylovanov and Zapechelnyuk (2013), Doornik (2014), Hörner, Morelli, and Squintani (2015), Zheng (2018), and Zheng and Kamranzadeh (2018).

Veto. In the event $\mathcal{V}$ player i receives (exogenous) payoff $V_i(\theta_i) \in (-\infty, 1]$.[7]

Escalation. In the event $\mathcal{E}$, players play a non-cooperative game. Its rules are a finite set of action profiles $\mathcal{A}$ and a mapping $(u_A, u_B) : \Theta^2 \times \mathcal{A} \to (-\infty, 1]^2$. We refer to the triple $(\mathcal{A}, u, \Theta^2)$ as the *escalation game.*

Settlement. Settlement, $\mathcal{Z}$, is an allocation $(x_A, x_B) \in [0, 1]^2$ with $x_A + x_B \leq 1$. In addition, player i receives a utility transfer $t_i \in \mathbb{R}$.

Payoffs. Player i type θ_i's ex-post payoff is equal to V_i, u_i, or x_i+t_i, depending on whether the final outcome is determined through veto, escalation, or settlement.

Solution Concept and Initial Information. Throughout our analysis we focus on perfect Bayesian equilibria (Fudenberg and Tirole, 1988). Everything but a player's type realization is commonly known ex-ante.

2.2 Arbitration Mechanisms.

We now provide an abstract description of an arbitration problem. In the next section, Section 3, we show that it is the consequence of five axioms and two assumptions that emerge naturally in arbitration settings and are common in the literature.

Arbitration. An arbitration problem is a choice problem among arbitration mechanisms.

An arbitration mechanism selects among the events $\mathcal{V}, \mathcal{Z}, \mathcal{E}$ *and* determines the outcome under settlement, $\mathcal{Z}$. Formally, we represent the set of all arbitration mechanisms via the set of admissible (reduced-form) direct revelation mechanisms (DRM). It consists of three parts: (i) an escalation probability γ; (ii) an expected settlement value z, and a signal Σ.[8]

Definition 1 (Reduced-Form Direct Revelation Mechanism)**.** A reduced-form DRM is a (collection of) mappings

$$\mathcal{M} = (\gamma, z, \Sigma) : \Theta^2 \to [0, 1] \times \mathbb{R}^2 \times \Delta(\mathcal{A}), \tag{$\mathcal{M}$}$$

where $\Delta(\mathcal{A})$ is the set of probability distributions over the action space $\mathcal{A}$.

[7] We show later that this includes cases in which payoffs in $\mathcal{V}$ are determined endogenously.

[8] In the next section, we show that this representation is without loss in canonical arbitration problems. We also provide further details on the three elements of the mechanism. For clarity we restrict ourselves here to the statements only.

Admissible reduced-form direct revelation mechanisms are *incentive compatible*, *individually rational*, and *satisfy the budget constraint*. For a formal description of these properties see page 12.

Timing. Players observe their types and an arbitration mechanism $\mathcal{M}$. They simultaneously decide whether to join the arbitration mechanism. If both agree to join, the mechanism is played and results in settlement $\mathcal{Z}$ or escalation $\mathcal{E}$. If one of them rejects the mechanism event $\mathcal{V}$ realizes. The events determine the final outcome.

3 A Canonical Arbitration Problem

In this section we formulate a set of axioms that constrain the designer of the arbitration mechanism. Our axioms arise naturally in real-world settings and are the unifying ingredients of the existing arbitration literature. Combined with two assumptions that guarantee scope for arbitration and a mild restriction on the designer's preferences these axioms form the basis of a canonical arbitration problem.

We then state and prove the applicable revelation principle that allows us to reduce the problem without loss to the mechanism defined in Section 2.

Our formulation constitutes a natural lower bound on the constraints an arbitrator faces. Constraining the arbitrator further is possible and sometimes done in the literature. We postpone a discussion of such modifications as well as the mapping from our setup to the existing literature to Section 5.

3.1 Basic Axioms

We formulate five economic axioms that are essential to various kinds of arbitration. We state the axioms first and discuss them in turns thereafter.

Axiom 1 (Types are ordered)**.** The functions u_i and V_i are increasing in θ_i.

Axiom 2 (Fights destroy surplus)**.** The joint payoff in the event $\mathcal{E}$ never exceeds 1, that is, for all type and action profiles in $\Theta^2 \times \mathcal{A}$ it holds that $u_A + u_B \leq 1$.

Axiom 3 (Unilateral Veto Rights)**.** An arbitration mechanism involves unilateral veto rights at the interim stage, that is, each player can veto the arbitration mechanism. A single veto triggers event $\mathcal{V}$.

Axiom 4 (Exogenous Rules of Conflict)**.** An arbitration mechanism treats $\mathcal{A}, u_i$, and V_i as exogenously given. Players can select any $a_i \in \mathcal{A}_i$ in event $\mathcal{E}$.

Axiom 5 (Budget Balance)**.** Arbitration cannot run a *structural deficit*, that is, any mechanism is budget balanced in expectation, $\sum_i \sum_{\theta_i} p(\theta_i)t_i(\theta_i) \leq 0$.

3.2 Discussion of the Axioms

Axiom 1 provides an interpretation of the player's type, θ_i: It is the privately known *strength in the fight.*[9]

Axiom 2 provides the economic rationale to why arbitration is desirable—to reduce the adverse selection problem. Given Axiom 2 there always is some settlement solution that can replicate the (ex-post) outcome of the fight. Thus, conditional on knowing the outcome of the fight, a settlement solution exists that is (weakly) preferred by both players to the fighting outcome.

Axiom 2 only concerns the relationship between $\mathcal{E}$ and $\mathcal{Z}$. To simplify the analysis, we take a reduced form approach on the veto payoff V_i and an ex-post notion is not well defined. Later we impose Assumption 1 (page 10) which implies that $\mathcal{V}$ is undesirable ex-post.

In some cases of arbitration a veto triggers a game too. Potentially the veto game is equivalent to the escalation game. We abstract from describing the game in our model description. In Proposition 3 we show that $V_i(\theta_i)$ captures the relevant aspects of a richer model that includes a veto game.

Jointly Axiom 1 and 2 provide the basis to a classic mechanism-design tradeoff. Strong players need to receive a high payoff under settlement to be willing to forgo the fight. A weak player may mimic a strong player to have access to that outcome but faces higher cost when ending up in the fight. Full efficiency requires full settlement.

Axiom 3 captures the idea that arbitration relies on mutual consent. Any player can enforce the event $\mathcal{V}$ by rejecting to participate in arbitration. The axiom *does not* exclude mandatory arbitration. For $V_i(\theta_i)$ low enough, participation is not an issue and the model is equivalent to the case in which *both* players are forced to participate. The purpose of Axiom 3 is to *allow* for veto rights.

Axiom 4 states that fights cannot be (completely) controlled by the designer. While escalation can be *selected* by the mechanism, the designer has no control over its *strategic environment.*

In reality the designer may have some influence on the strategic environment. The arbitrator's choice among different potential games can easily be incorporated

[9] We show below (Proposition 5) that we can also interpret it as the *value of winning.* However, in some cases the problem becomes trivial under that interpretation.

in the formulation of u_i and $\mathcal{A}$. Yet under Axiom 2 all such solutions are inefficient. Implicitly our formulation assumes that the escalation game $(\mathcal{A}, u_i, \Theta^2)$ is the *best* among those available from the designer's perspective. To best of our knowledge all of the literature on arbitration prespecifies the escalation game.

Finally, Axiom 5 concerns the arbitrator's budget. It states the designer cannot trivially overcome the adverse selection problem by adding an arbitrary amount of money to the system. We take the mildest form of a budget constraint, by assuming that budget balance has to hold at an ex-ante stage. Stronger versions are not more restrictive due to risk-neutrality of the players (see the discussion at the end of this section).

Jointly Axiom 3 and 4 capture the idea that arbitration can only control part of the environment. While arbitration has full control over settlement, it has limited control over the fight.

3.3 Primitives to the Arbitrator and Basic Assumptions

To discuss optimal arbitration we need to define the scope of it. We do so by means of two mild assumptions. It is instructive to impose them at the level of the *primitives to the arbitrator's problem.* While primitives to the design problem, these terms may be derived from the model primitives stated in Section 2.

The primitives to the arbitrator's problem are $\{\Theta, p, U_i, V_i\}$. All but U_i, the expected payoff from escalation, are part of our model primitives. The function U_i describes a reduced-form formulation of the outcome in $\mathcal{E}$. The function U_i is a primitive to the arbitration problem. The rules of conflict are exogenous and the arbitrator can at most influence information sets, but not the shape of the continuation-utility functions. We now derive U_i.

Deriving U_i. The expected utility U_i depends on the triple $(\sigma_i, \theta_i, \mathcal{B})$ that describes i's information set at the beginning of the escalation game. Specifically, $\mathcal{B}$ is the public information structure and (σ_i, θ_i) is i's additional private information.

At each decision node a player best responds to her current information set. A player's information set consists of a private part and a public part. Player i's private part, (σ_i, θ_i), is her exogenous private information, θ_i, and all further information privately acquired previously, σ_i. The public part, the information structure $\mathcal{B}$, contains all elements that are public knowledge at the decision node.

Assume we are at a point of the game in which the mechanism has terminated and event $\mathcal{E}$ realized. Player i uses her information set to form beliefs. She computes a probability mass function over the opponent's payoff type, $\beta_i : \Theta \to$

$[0,1]$—her belief about the opponent's type. In addition, the player forms a set of conditional probability mass functions $f_{\theta_{-i}} : A_{-i} \to [0,1]$—her belief about θ_{-i}'s continuation strategies. Given her beliefs, the player picks an action that constitutes a best response to these beliefs.

On the equilibrium path players' (conditional) beliefs are correct. The expected continuation value for the (on-path) event $\mathcal{E}$ is

$$U_i(\sigma_i, \theta_i, \mathcal{B}) := \max_{a_i \in A_i} \sum_{\theta_{-i}} \beta_i(\theta_{-i}|\sigma_i, \mathcal{B}) \sum_{A_{-i}} f_{\theta_{-i}}(a_{-i}|\sigma_i, \mathcal{B}) u_i(\theta_i, \theta_{-i}, a_i, a_{-i}), \tag{U}$$

and the function $\beta_i(\theta_i|\cdot)$, $f_{\theta_{-i}}(a_{-i}|\cdot)$ correspond to the correct (on-path) beliefs about the opponent's type and her associated action choices.[10]

Treating the function U_i as a primitive we make one of the following two implicit assumptions. Either (i) the equilibrium selection in the continuation game is exogenous and known to the arbitrator or (ii) the arbitrator can control the equilibrium selection and picks the "best" equilibrium given her objective.

Either of these assumptions implies that we focus on implementability of the optimal mechanism. Approaches in which the arbitrator is unaware of the equilibrium-selection rule are beyond the scope of our paper.

Basic Assumptions. We make two additional assumptions. The first ensures that there is room for arbitration, the second determines the range of arbitration objectives.

Assumption 1. $U_i(\check{\sigma}_i, \theta_i, \check{\mathcal{B}}) \geq V_i(\theta_i)$ for $\{\check{\sigma}_A, \check{\sigma}_B, \check{\mathcal{B}}\}$ such that $\beta_i(\theta_{-i}|\check{\sigma}_i, \check{\mathcal{B}}) = p(\theta_{-i})$ and $\sigma_i = \emptyset$ for each player.

Assumption 1 implies that an arbitration mechanism with full participation in equilibrium exists. The arbitrator escalates all cases and players receive at least their veto payoffs. Assumption 1 excludes cases in which parties are *exogenously punished* for participation.

Assumption 1 captures cases in which participation in arbitration is neutral or beneficial even if it fails to settle. Benefits may come through psychological components or institutional features such as penalties for not agreeing to arbitration. Examples can be found both in legal systems through explicit sanctions imposed by the court and in international relations where participation in peace talks often implies a temporary lift of imposed sanctions.

Our second assumption concerns the designer's preferences.

[10] The commonly known $\mathcal{B}$ includes the belief closed subset under which the game is played. Players have (some) common prior and there is common certainty of rationality (see Bergemann and Morris, 2016, for further details).

Assumption 2. The arbitrator's preferences are weakly monotone in cases settled and in player's aggregate utility. In particular, suppose the following is true for two mechanisms $\mathcal{M}$ and $\mathcal{M}'$.

1. $\mathcal{M}'$ settles at least as many cases as $\mathcal{M}$;
2. the (ex-ante) expected payoff of either player in mechanism $\mathcal{M}'$ is at least as large as her payoff in mechanism $\mathcal{M}$.

Then the arbitrator (weakly) prefers $\mathcal{M}'$ to $\mathcal{M}$.

Potential objective functions under Assumption 2 include *minimize the ex-ante probability of conflict* and *maximize the player's ex-ante expected utilities*, the objectives mainly used in the literature.

Combining Assumption 1 and 2 and Axiom 1 to 5 defines a canonical arbitration problem.

Definition 2 (Canonical Arbitration Problem)**.** A canonical arbitration problem is a mechanism-design problem under Axiom 1 to 5 and Assumption 1 and 2.

3.4 A Revelation Principle for Optimal Arbitration

We now state and prove a revelation principle for optimal arbitration. We start by defining a direct revelation mechanism which we then reduce further to the *reduced-form* DRM from Section 2. Our revelation principle states that it is without loss to focus on reduced-form DRM.

A direct revelation mechanism maps the profile of type reports into four objects: (i) a probability, γ, that the conflict escalates to event $\mathcal{E}$, (ii) a sharing rule, X, determining the allocation when settlement is achieved, (iii) a direct utility transfer t, and (iv) an additional signal, random variable Σ, with realization (σ_A, σ_B). Without loss we assume that σ_i, is an action recommendation in the escalation game.

Definition 3 (Direct Revelation Mechanism)**.** A direct revelation mechanism (DRM) is a mapping

$$\mathcal{M}^e(\cdot) = (\gamma(\cdot), X(\cdot), t(\cdot), \Sigma(\cdot)) : \Theta^2 \to [0,1] \times [0,1]^2 \times \mathbb{R}^2 \times \Delta(\mathcal{A}), \qquad (\mathcal{M}^e)$$

In a DRM $\mathcal{M}^e$ players report their type privately to the mechanism. With some probability γ the event $\mathcal{E}$ follows and a signal realization σ_i is privately communicated to player i. With the remaining probability $1-\gamma$ the event $\mathcal{Z}$ follows and an allocation (x_A, x_B) is implemented. In addition, players receive utility transfers according to t. Parties are risk neutral and it is without loss to assume transfers are only paid in event $\mathcal{Z}$.

Conditional on participation and truthful reporting by $-i$, the expected payoff in a given DRM depends on the type report m_i and the actual type θ_i. It is

$$\begin{aligned}\Pi_i(m_i;\theta_i) =& \underbrace{\sum_{\theta_{-i}} p(\theta_{-i})(1-\gamma_i(m_i,\theta_{-i}))(x_i(m_i,\theta_{-i})+t_i(m_i,\theta_{-i}))}_{=:z_i(m_i)(\text{settlement value})} \\ &+ \underbrace{\left(\sum_{\theta_{-i}} p(\theta_{-i})\gamma_i(m_i,\theta_{-i})\right)\sum_{\sigma_i} Pr(\sigma_i|m_i)U_i(\sigma_i;\theta_i,\mathcal{B})}_{=:y_i(m_i;\theta_i)(\text{escalation value})},\end{aligned} \tag{1}$$

with $Pr(\sigma_i|m_i)$ the (conditional) probability that a signal σ_i realizes. The first part, the settlement value, is the expected payoff from event $\mathcal{Z}$. It depends only on the type report m_i. The second part, the escalation value, is the expected payoff from event $\mathcal{E}$. It depends on the type report m_i and the actual type θ_i.

A DRM is *incentive compatible* if

$$\forall m_i, \theta_i \in \Theta: \quad \Pi_i(\theta_i;\theta_i) \geq \Pi_i(m_i;\theta_i).$$

A DRM is *incentive feasible* if it is incentive compatible and participation is individually rational, that is,

$$\forall \theta_i \in \Theta: \quad \Pi_i(\theta_i;\theta_i) \geq V_i(\theta_i).$$

A DRM is *admissible* if it is incentive feasible and satisfies the arbitrator's budget constraint, that is,

$$\sum_i \sum_{\theta_i} p(\theta_i) \sum_{\theta_{-i}} p(\theta_{-i})(1-\gamma_i(\theta_i,\theta_{-i}))t_i(\theta_i,\theta_{-i}) \leq 0.$$

Since x_i and t_i are tedious to formulate, it is helpful to reduce the choices to *expected* terms. A reduced-form DRM captures exactly that. Reduced-form DRM and DRM are equivalent by reformulating the arbitrator's budget constraint.

Lemma 1. *A reduced-form DRM (γ, z, Σ) is admissible if and only if an incentive feasible DRM (γ, X, t, Σ) exists that satisfies*

$$\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \leq 1 - \sum_{(\theta_A,\theta_B)} p(\theta_A)p(\theta_B)\gamma(\theta_A,\theta_B). \tag{BB}$$

Proposition 1 (Revelation Principle)**.** *It is without loss of generality to focus on admissible reduced-form DRMs.*

An immediate corollary to Proposition 1 is reminiscent of the findings of Compte and Jehiel (2009) and Zheng (2018).

Corollary 1. *Full settlement, $\gamma(\cdot) = 0$, is admissible if and only if $V_A(K) + V_B(K) \leq 1$.*

Discussion. Before proceeding with the analysis, we briefly comment on some of our assumptions not implied by our axioms. Relaxing players' commitment to obey the arbitrator, ex-ante symmetry, and type independence do not alter any of our results. However, the relaxations are notationally inconvenient. Our model is identical to one in which budget balance holds at an ex-post level as players are risk neutral (see the arguments in Börgers and Norman, 2009).

While u_i allows *interdependencies*, we exclude *correlation* between the players' type distributions. Correlation would—provided a full-rank condition—free the arbitrator entirely from incentive constraints precisely as in Crémer and McLean (1988).[11]

4 Belief Management

In this section we derive an information-design problem conditional on the escalation game. Solving that problem is necessary and sufficient to find optimal arbitration if full settlement cannot be guaranteed. The necessary part follows from Axiom 4. We show that finding the optimal information structure is sufficient to determine the optimal mechanism too. We focus on cases in which a mechanism implementing full settlement does not exist.[12]

Assumption 3 (No full settlement)**.** $V_A(K) + V_B(K) > 1$.

We derive the result in steps. First, we establish our notion of an information structure. Second, we characterize a mapping from implementable information structures to a unique candidate for the optimal mechanism. Third, we apply

[11] In an extension we consider the case without transfers. In that case the designer faces an additional constraint, equation (GI) defined on page 28. In these cases correlated types are possible and constraint equation (GI) holds with equality.

[12] Assumption 3 is useful to rule out the trivial full-settlement solution when considering a symmetric (i.e., anonymous) objective. For the sake of simplicity, we build our argument ignoring cases in which full settlement is admissible but not optimal. However, nothing changes qualitatively when including those cases expect that the last step of the proof of Theorem 1 becomes slightly more involved.

this result to a range of objectives of the arbitrator. We obtain the corresponding information-design problem.

Our results simplify the arbitration problem. Given an information structure in $\mathcal{E}$ the optimal mechanism is the solution to a system of linear equations. The observations leading to Theorem 1 characterize these equations. Through Theorem 1 we can disentangle the mechanism-design part of eliciting information from the information-design part of distributing that information.

The mapping from information structures to the optimal mechanism characterizes the cost of implementing a certain information structure. The information-design problem of finding the optimal information structure has an intuitive economic interpretation and allows us to describe how properties of the escalation game influence the optimal mechanism.

4.1 Preliminaries: Information Sets and Notation

The information set at the beginning of the escalation stage consists of a private part and a public part. The private part, (σ_i, θ_i), contains the privately observed realization of the signal, a player's own private history of play, and the exogenous private type. The public part contains everything commonly known, in particular, the arbitrator's choice $\mathcal{M}$, the equilibrium strategies, the information that the conflict escalated, and the common prior.

We derive a formulation of the commonly known information structure at the beginning of the escalation game. For these purposes it is useful to assume that the mechanism first publicly announces escalation, and thereafter privately communicates the realizations (σ_A, σ_B) of signal Σ to the players.

We begin by describing the information structure conditional on the public announcement of the realization of event $\mathcal{E}$. The probability $\beta_i(\theta_j|\theta_i)$ is the probability that a player i who reported to be type θ_i attaches to the event that player j has reported to be type θ_j. The probability distribution $\beta_i(\cdot|\theta_i)$ collects all such beliefs.

Because γ is a function of the report profile, the distributions depend on a player's own report. Because γ is commonly known, all on-path probability distributions $\beta_i(\cdot|\theta_i)$ are commonly known too. We collect them in the set B. The set B is sufficient to describe the information structure *net of signal* Σ.

The signal Σ provides additional information to the players. While the beliefs B are common from second order onwards, Σ can potentially induce non-common higher-order beliefs. For our purposes, it is useful to keep the function Σ as a sufficient statistics for the induced higher-order beliefs.

We represent the public information structure as $\mathcal{B} := (B, \Sigma)$. In appendix E we provide a detailed discussion on the updating procedure and the information structure. Here we focus on its properties. We highlight which information structures can be induced, what is a minimal description of an information structure, and what is its purpose in the escalation game.

Definition 4. A mechanism $\mathcal{M}$ *induces* information structure $\mathcal{B}$ if $\mathcal{B}$ is the public information structure in some on-path continuation game of event $\mathcal{E}$.

Definition 5 (Consistency)**.** A set of beliefs over the type space, B, is *consistent* with respect to the prior, p, if there is a mechanism $\mathcal{M}$ and a set S of beliefs about signal realizations such that $\mathcal{M}$ induces (B, S).

In the next lemma we exploit that consistency pins down some relationship between the different β_i through Bayes' rule. If beliefs are probability distributions with full support over Θ, an equivalent requirement to consistency of B is to determine an arbitrary subset $\{\beta_A(\cdot|m)\}_{m\in\Theta} \cup \beta_B(\cdot|1)$. Bayes' rule determines the remaining beliefs.

Lemma 2. *Take a set of arbitrary probability mass functions each with full support over* Θ*,* $\{\beta_A(\cdot|m)\}_{m\in\Theta} \cup \beta_B(\cdot|1)$*. There is a unique set* $\{\beta_B(\cdot|m)\}_{m\in\Theta\setminus\{1\}}$ *such that* $B = \{\beta_i(\cdot|m)\}_{m\in\Theta, i\in\{A,B\}}$ *is consistent.*

Any B that can realize in any continuation game can be introduced directly by some γ.

Lemma 3. *B is consistent if and only if there is a γ such that B follows from applying Bayes' rule under γ.*

We combine the two lemmas above to state a result that allows us to connect our notion of a public information structure to the information-design literature.

Lemma 4. *Any information structure can be represented by a tuple* (B, Σ)*. Moreover, a mechanism exists that induces information structure* $\mathcal{B} = (B, \Sigma)$ *if and only if B is consistent.*

The last result can be interpreted as follows. The continuation game in the event $\mathcal{E}$ is an *incomplete information game* in the sense of Bergemann and Morris (2016). Using their terminology, a basic game consist of four elements. An action set, $\mathcal{A}$, a payoff function, u, a type space, Θ^2, and a common prior, B. An incomplete information game is a basic game augmented by a random variable, Σ, determining further, privately held, information.

The first-order and second-order beliefs about the state that players hold when observing the mechanism's outcome are entirely determined by the escalation rule γ as the next corollary to Lemma 3 and 4 shows.

Corollary 2. *A reduced-form DRM, (z, γ, Σ), induces information structure, (B, Σ) if and only if B follows from applying Bayes' rule under γ.*

4.2 Belief Management

In the previous part we have shown that γ provides the "prior" B to the incomplete information game $(\mathcal{A}, u, \Theta^2, B, \Sigma)$. In this part we show that (B, Σ) is sufficient to determine a unique candidate mechanism (z, γ, Σ) in a canonical arbitration problem. We construct a mapping $M_\Sigma(B) \mapsto (z, \gamma)$. It determines the arbitrator's least-costly option to induce (B, Σ). It is the solution to a system of linear equations.

Definition 6. An information structures is *implementable* if an admissible reduced-form DRM induces it.

Theorem 1. *Consider a canonical arbitration problem. The set of implementable information structures (B, Σ) is compact. Moreover, for any implementable (B, Σ) the optimal admissible reduced-form mechanism, (z, γ, Σ), is unique.*

Theorem 1 is our main result. The proof is constructive and characterizes the linear equations that determine the function M_Σ. We organize our discussion of the intuition using a set of observations that correspond to the steps in the formal proof. Recall that B is formed immediately after event $\mathcal{E}$ realized but *before* the signal realizations (σ_A, σ_B) have been communicated.

Observation 1. Both the belief about the state, B, and the continuation payoffs, U_i, are homogeneous of degree 0 in the escalation rule. The escalation value, y_i, is homogeneous of degree 1 in the escalation rule.

Suppose player i submits a report m_i and learns about the event $\mathcal{E}$. Before receiving any additional signal, the probability of facing a particular type $\tilde{\theta}_{-i}$ is

$$\beta_i(\tilde{\theta}_{-i}|m_i) = \frac{p(\tilde{\theta}_{-i})\gamma(m_i, \tilde{\theta}_{-i})}{\sum_{\theta_{-i}} p(\theta_{-i})\gamma(m_i, \theta_{-i})}.$$

That probability is determined by the *relative* likelihood of escalation only. Thus, if γ implies B so does $\alpha\gamma$. The effect of γ on the continuation payoff in $\mathcal{E}$ is entirely expressed via B. Thus, any U_i is invariant to any scaling of γ. Finally, the (interim) probability of reaching escalation and hence the escalation value are linear in γ, realizations σ are constant in γ.

Observation 2. The *most-costly escalation rule* inducing B is unique.

Take any γ that induces B and pick the largest scalar $\overline{\alpha}$ such that $\overline{\alpha}\gamma(\theta_A, \theta_B) \leq 1$ for all (θ_A, θ_B). Then, the rule $\overline{g}_B := \overline{\alpha}\gamma$ minimizes the right-hand side of the budget constraint (BB). It is the most-costly rule for the arbitrator. Identifying the most-costly escalation rule is sufficient to characterize *all escalation rules* that induce B. The set of all γ inducing B is $\{\alpha\overline{g}_B : \alpha \in (0, 1]\}$. Given B, the problem reduces to finding the *lowest* α such that (B, Σ) is implementable.

Our next observation contains the main step towards the result.

Observation 3. It is without loss to assume that any type θ_i has either a binding incentive constraint or a binding participation constraint. Given (B, Σ), all constraints are linear in α and z.

Assumption 2 and 3 ensure that some constraint binds for any type. Otherwise a Pareto improving mechanism with less escalation exists. The second part follows by combining players' expected payoffs with Observation 2.

Observation 3 implies that (B, Σ) captures the entire non-linear part of the constraints. Given (B, Σ) each type has some binding constraint and the set of constraints consists of $2K$ independent linear equations. We have $2K + 1$ unknowns, the $2K$ settlement values and the scalar α. To close the problem we need one more equation. We use the arbitrator's budget constraint (BB).

Observation 4. It is without loss to assume that (BB) holds with equality.

By Observation 3, z_i is linear in α, and thus $\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i)$ is linear too. The right-hand side is independent of z and linear in α. Solving for α delivers a unique tuple $(z^*_\Sigma, \alpha^*_\Sigma \overline{g}_B)$ satisfying the binding constraints with equality.

Via Observation 1 and 2, an admissible escalation rule inducing B exists if and only if the corresponding $\alpha^*_\Sigma \leq 1$.

We construct a function $M_\Sigma : B \mapsto (z, \gamma)$ that (given Σ) identifies a *unique candidate* (z, γ) for any implementable B that can be induced by an admissible $\mathcal{M}$. It points to the origin otherwise. The function is given by

$$M_\Sigma(B) := \mathbb{1}_{\alpha^*_\Sigma \leq 1} \Big(z^*_\Sigma, \alpha^*_\Sigma \overline{g}_B \Big),$$

and continuous in the interior of its support.

Discussion of Theorem 1. The canonical arbitration problem contains a mechanism-design part and an information-design part. The arbitrator has full control over the settlement environment, yet under escalation her power ceases and players are free to take decisions.

The informational content of sending players to escalation is endogenous. Moreover, the arbitrator has to evaluate the effects of information revelation on escalation. Thus, it is necessary that the arbitrator solves an information-design problem for that event.

Theorem 1 implies that solving that information-design problem is sufficient too. Given a "prior", B, and a signal structure, Σ, the optimal mechanism is pinned down by a set of linear equations.

Note that the arbitrator has more power in the information-design problem implied by Theorem 1 than in Bergemann and Morris (2016). Under Theorem 1, the arbitrator can produce the "prior" to her information-design problem at a cost. The literature on information design assumes that the prior is exogenously given. The reason for that difference is precisely that the arbitrator can control the strategic environment prior to escalation.

The main implication of Theorem 1 is that we can formulate the entire problem focusing only on the event $\mathcal{E}$. That is, all constraints and the objective are a function of the information structure (B, Σ) only. The set of functions M_Σ determines if such an information structure is implementable and how it is best implemented.

We want to emphasize that the (re-)formulation simplifies the analysis in several ways. Keep in mind that for given "prior", B, the information-design part of finding the optimal Σ is necessary for any formulation of the arbitration problem. We leave it unchanged in our formulation.

Before this stage our formulation simplifies the problem. On the level of finding a solution, Theorem 1 not only characterizes the set of *implementable* B that some admissible $\mathcal{M}$ can induce, but also limits it to a single candidate mechanism for any B.

Our simplification is independent of the objective itself. The only requirement we impose is the monotonicity from Assumption 2. Theorem 1 implies that B is a sufficient statistics for a candidate mechanism for any objective satisfying Assumption 2. The function $M_\Sigma(B)$ is continuous on the interior of its support. It provides an intuitive mapping from implementable Bs to mechanisms, driven by the expected settlement shares necessary to support B.

Focusing on the event $\mathcal{E}$ emphasizes that the key to arbitration is what happens if arbitration fails. It illustrates that the properties of the escalation game matter and need to be understood for successful arbitration. For a given escalation game

Theorem 1 links the properties of the mechanism itself directly to the information it induces. The information-design approach allows us to make predictions on the information arbitration reveals.

Our next step is to characterize the information-design problem for a class of objective functions. The characterization provides an intuitive description of the main trade-offs and delivers insights even without specifying the escalation game. Moreover, for a *given escalation* game the solution can be directly computed.

4.3 Optimal Belief Management

In this part we apply Theorem 1 to a class of objective functions of the arbitrator. Our formulation covers a range of social welfare functions. It includes all cases in which social welfare is players' aggregate utility minus some additional cost that escalating the conflict imposes on society. Our result states a well-defined problem of selecting an information structure $\mathcal{B}$.

We refer to the problem of finding $\mathcal{B}$ as the *belief-management* problem. It is an economically intuitive, compact representation of the arbitration problem, and entirely based on the escalation game $(\mathcal{A}, u, \Theta^2)$.

We start by stating a class of objective functions. Take any $\xi \in [0,1]$, and assume the arbitrator chooses an admissible mechanism (γ, z, Σ) that solves

$$\max_{(\gamma,z,\Sigma)} (1-\xi) \underbrace{\left(\sum_i \mathbb{E}[z_i(\theta) + \Big(\sum_{\theta_{-i}} p(\theta_{-i})\gamma_i(\theta,\theta_{-i})\Big)\widehat{U}_i(\theta;\theta,\mathcal{B})]\right)}_{\text{expected joint surplus}} - \xi \underbrace{Pr(\mathcal{E})}_{\text{prob. of } \mathcal{E}}, \quad (P_{\mathcal{M}})$$

with

$$\widehat{U}(m_i;\theta_i,\mathcal{B}) := \sum_{\sigma_i} Pr(\sigma_i|m_i) U_i(\sigma_i;\theta_i,\mathcal{B}).$$

Problem $(P_{\mathcal{M}})$ covers any convex combination between conflict minimization ($\xi = 1$) and joint surplus maximization ($\xi = 0$).

Constraints. A mechanism is admissible if it satisfies participation constraints, incentive constraints and is budget balanced. Theorem 1 states that instead of an admissible mechanism we can directly choose an information structure, $\mathcal{B} = (B, \Sigma)$, such that B is consistent and in the support of $M_\Sigma(B)$. What is left is to determine how to find the *optimal* $\mathcal{B}$.

To facilitate intuition, think of Σ as (optimally) chosen given prior B. If Σ is optimal conditional on B, the arbitrator's problem is to implement B. However, B not only influences the choice of Σ, but also determines the marginal type distributions conditional on $\mathcal{E}$, denoted by $(\rho_A(\cdot), \rho_B(\cdot))$. The distributions ρ_i are the solution to the following system of equations determined by B[13]

$$\rho_i(\theta_i) = \sum_{\theta_{-i}} \beta_{-i}(\theta_i|\theta_{-i})\rho_{-i}(\theta_{-i}) \qquad \forall\theta_i, \theta_{-i}. \tag{2}$$

Theorem 1 incorporates one binding constraint per type and budget balance in $M_\Sigma(B)$. What remains is to state the remaining incentive constraints as a funciton of $\mathcal{B}$. Take any i and $\theta, \theta' \in \Theta$. Incentive compatibility holds if and only if[14]

$$\frac{\rho_i(\theta')}{p(\theta')}\left(\widehat{U}_i(\theta';\theta',\mathcal{B}) - \widehat{U}_i(\theta';\theta,\mathcal{B})\right) - \frac{\rho(\theta)}{p(\theta)}\left(\widehat{U}_i(\theta;\theta',\mathcal{B}) - \widehat{U}_i(\theta;\theta,\mathcal{B})\right) \geq 0. \tag{IC}$$

Objective. To facilitate the exposition we impose more structure on the problem. We impose that apart from the strongest type all types have a (pre-arbitration) incentive to seek settlement.

Assumption 4 (Close Conflicts)**.** $\sum_i \sum_{\theta_i \in \hat{Q}} p(\theta_i)V_i(\theta_i) < \sum_{\theta_i \in \hat{Q}} p(\theta_i)$, for any $\hat{Q} \subseteq \Theta$ and $\hat{Q} \neq \{K\}$.

Assumption 4 imposes structure on the set of relevant constraints.[15]

Lemma 5. *Under Assumption 1 to 4 the strongest type's participation constraint holds with equality. All other participation constraints are redundant.*

Jointly, Observation 3 and Lemma 5 imply that we can identify a path, $\iota_i : \Theta \setminus K \to \Theta$, such that $\Pi_i(\iota_i(\theta_i), \theta_i) = \Pi_i(\theta_i; \theta_i)$, and $\iota_i(\theta) \neq \theta$. The function ι_i determines a binding incentive constraint for each type but the strongest for whom the participation constraint binds by Assumption 3. For the special case that local constraints are sufficient for incentive compatibility $\iota_i(\theta_i) = \theta_i{+}1$. Given $\iota_i(\theta_i)$ we can define $D_i^\iota(m_i; \theta_i, \mathcal{B}) := \widehat{U}_i(m_i; \iota_i(\theta_i), \mathcal{B}) - \widehat{U}_i(m_i; \theta_i, \mathcal{B})$. We refer to $D_i^\iota(m_i, \theta_i, \mathcal{B})$ as θ_i's *type disadvantage* as it describes θ_i's loss in utility compared to the "next best" type from θ_i's perspective.

We can use D_i^ι to determine a type's *virtual loss.* Let $\Theta^\iota(\theta) := \{k \in \Theta | \exists n \geq 0 \text{ s.t. } \iota^{\circ n}(k) = \theta\}$ be the set of types on path ι that lead to θ after some iteration $\iota^{\circ n}$. Define

$$w_i^\iota(\theta_i) := \sum_{k \in \Theta^\iota(\theta)} \frac{p(k)}{p(\iota(\theta_i))}.$$

[13] The proof of Lemma 2 provides the relevant arguments.

[14] For details, see proof of Theorem 1 in particular step 3 an the Lagrangian in appendix B.

[15] We derive a general version absent Assumption 4 in the appendix to the paper. The main difference is that it complicates identifying the set of binding constraint.

For the special case that local incentive constraints bind, $\iota(\theta) = \theta + 1$, ω^ι describes the hazard rate.

Definition 7 (Virtual Loss)**.** Player θ_i's virtual loss of pretending

$$\widehat{\Psi}_i^\iota(\theta_i, \mathcal{B}) := \begin{cases} w_i^\iota(\theta_i) D_i^\iota(\iota_i(\theta_i); \theta_i, \mathcal{B}) & \text{if } \theta_i \neq K \\ 0 & \text{otherwise.} \end{cases}$$

We state the objective as a function of two objects, each defined within $\mathcal{E}$.

$$\begin{aligned} \mathbb{E}[\widehat{\Psi}|\mathcal{B}] &:= \sum_i \sum_{\theta_i=1}^{K-1} \rho_i(\iota_i(\theta_i)) \widehat{\Psi}_i(\theta_i, \mathcal{B}) \quad \text{(expected virtual loss)} \\ \mathbb{E}[\widehat{U}|\mathcal{B}] &:= \sum_i \sum_{\theta_i=1}^{K} \rho_i(\theta_i) \hat{U}_i(\theta_i; \theta_i, \mathcal{B}) \quad \text{(expected utility)} \end{aligned} \tag{3}$$

In principle, multiple incentive constraints could bind. Let $(IC)^A$ be the set of all incentive constraints as defined in equation (IC) which are not governed by ι_i. Define the optimization problem[16]

$$\min_{\mathcal{B}} \frac{\xi + (1-\xi)\left(1 - \mathbb{E}[\widehat{U}|\mathcal{B}]\right)}{\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}] - 1}, \text{ s.t. } (IC)^A \text{ and } M_\Sigma(B) \neq 0. \tag{$P_\mathcal{B}$}$$

Proposition 2 (Duality)**.** *Take a canonical arbitration problem under Assumption 3 and 4. A mechanism solves $(P_\mathcal{M})$ if and only if $(\gamma, z) = M_{\Sigma^*}(B^*)$ and (B^*, Σ^*) solves $(P_\mathcal{B})$.*

Proposition 2 follows from using the function $M_\Sigma(B)$ to replace z and γ in the arbitrator's objective and rearranging terms. Recall that a consistent B follows from any arbitrary $\{\beta_A(\cdot|j)\}_{j\in\Theta} \cup \beta_B(\cdot|1)$ under Lemma 2.

Proposition 2 provides an intuitive belief-management problem which is equivalent to the canonical arbitration problem. Before discussing the general formulation it is useful to consider the two polar cases. First, if the arbitrator maximizes joint surplus ($\xi = 0$), she seeks to *maximize*

$$\frac{\mathbb{E}[\widehat{\Psi}|\mathcal{B}]}{1 - \mathbb{E}[\widehat{U}|\mathcal{B}]}.$$

Second, if the arbitrator minimizes escalation ($\xi = 1$), she seeks to *maximize*

$$\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}].$$

[16] Note that $\mathbb{E}[\widehat{\Psi}|\mathcal{B}]+\mathbb{E}[\widehat{U}|\mathcal{B}] > 1$ for any information structure with $M_\Sigma(B) \neq 0$ by Theorem 1.

The case $\xi = 1$ has an analogue in revenue-maximizing auction design (Myerson, 1981). The main difference is that—although types are ordered—the term $\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}]$ is non-linear in the arbitrator's choice. These non-linearities increase complexity. However, conceptually an arbitrator minimizes the likelihood of the conflict by maximizing the expected virtual valuation of the escalation game over the information structure.

More generally, the arbitrator wants to decrease the numerator of the objective in $(P_{\mathcal{B}})$. That term captures the cost of escalation to the arbitrator. They consist of a fixed component, ξ, and a variable component, $(1-\xi)(1-\mathbb{E}[\widehat{U}|\mathcal{B}])$, that captures the (joint) surplus loss of escalation. The cost of escalation are, however, only conditional on the event $\mathcal{E}$. Therefore, we have to multiply the cost by $(\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}] - 1)^{-1}$. The lower the likelihood of escalation, the lower the expected cost. Thus, the arbitrator wants to maximize the denominator to ensure settlement as often as possible.

We now address the two terms separately. The intuition for the numerator's form is straight-forward. The higher the welfare losses from escalation, and the more intense the arbitrator's preferences about it, the more costly is escalation.

The intuition for the denominator's form is more subtle. To minimize the likelihood of escalation, the arbitrator increases the (expected) virtual loss and utility for players in the event $\mathcal{E}$. Consider a player in event $\mathcal{E}$. Increasing her virtual loss contributes to satisfying incentive constraints. The higher $\mathbb{E}[\widehat{\Psi}|\mathcal{B}]$, the lower the information rent the arbitrator has to pay to ensure incentive compatibility. Increasing players' utility, in turn, incentivizes players to agree to participate in arbitration in the first place. It relaxes participation constraints.

Proposition 2 characterizes the economic forces. It provides an intuition how continuation play affects the arbitrator's choices. In the next section we address it in detail.

Complexity increases when optimizing the choice of the signal Σ. However, combining Proposition 2 and Theorem 1 implies that the complexity is a direct consequence of the associated information-design problem in arbitration. That is, if we restrict—as most of the literature—complexity such that the scope for information design in the game $(\mathcal{A}, u, \Theta^2)$ is tractable, so is the arbitration problem. Our results also show, however, that restricting the influence of information on outcomes in the escalation game is not without loss. In fact, the effect of information on outcomes is precisely what drives optimal arbitration.

We discuss some avenues on solving the information-design problem in Section 5.

4.4 The Economics of Optimal Arbitration

Proposition 2 highlights the fundamental trade-off of optimal arbitration: Balancing welfare maximization in the escalation game, $\mathbb{E}[\hat{U}|\mathcal{B}]$, against maximizing discrimination, $\mathbb{E}[\hat{\Psi}|\mathcal{B}]$. Full settlement is impossible if the designer has to offer strong types large shares to make them participate. Full settlement fails because large shares incentivize weak types to mimic the strong. The more discriminatory the escalation game, the lower the incentives for weak types to mimic the strong. Consequently, the more effective the screening which, in turn, allows for larger settlement shares granted to the strong types.

This trade-off between incentives to participate and deterrence of mimicking behavior is robust to the arbitrator's preferences (i.e., ξ). Even an arbitrator who only cares about cases settled ($\xi = 1$) sacrifices some discrimination to raise the players' welfare conditional on escalation. In reverse, an arbitrator who only cares about the players' joint surplus ($\xi = 1$) sacrifices some welfare to discriminate more effectively.

Surprisingly, most of the models in the literature shut down the welfare channel and focuses on the case where arbitration is only about the effectiveness of screening (see Bester and Wärneryd, 2006; Doornik, 2014; Hörner, Morelli, and Squintani, 2015, among others).

A simple corollary to Proposition 2 describes the condition under which welfare is absent in the designer's consideration.

Corollary 3. *Suppose* $\mathbb{E}[\hat{U}|\mathcal{B}]$ *is constant in* $\mathcal{B}$. *Then an arbitrator, for any* ξ, *wants to maximize discrimination* $\mathbb{E}[\hat{\Psi}|\mathcal{B}]$.

The literature referred to above makes the assumption that the joint surplus in the event $\mathcal{E}$ is reduced to a number $E < 1$. The type-profile determines only *how* E is distributed. It is immediate from the corollary that in such cases the maximization of joint surplus and the minimization of escalation coincide.

The following specification describes a (binary) version of the case referred to above.[17] Assume $\Theta = \{1, 2\}$ and $U_i(m_i; 1, \mathcal{B}) = \beta_i(1|m_i)(E/2) + \beta_i(2|m_i)(E/3)$ and $U_i(m_i; 2, \mathcal{B}) = \beta_i(1|m_i)(2E/3) + \beta_i(2|m_i)(E/2)$. In this case, we get a single type disadvantage

$$D_i(2, 2, \mathcal{B}) = U_i(2; 2, \mathcal{B}) - U_i(2; 1, \mathcal{B}) = \frac{E}{6}\beta_i(1|2) - \frac{E}{6}\beta_i(2|2).$$

[17] This formulation is equivalent to the model of Hörner, Morelli, and Squintani, 2015. We revisit (a superset of) this example in Section 5.1.

The arbitrator maximizes

$$\mathbb{E}[\hat{\Psi}|\mathcal{B}] = \sum_i \rho_i(2)\frac{1-p(2)}{p(2)}D_i(2,2,\mathcal{B}) = \frac{1-p(2)}{p(2)}\frac{E}{6}\sum_i \rho_i(2).$$

Whether $\rho_i(2)$ is implementable depends on the designer's budget constraint (i.e. if $\alpha \leq 1$). The designer picks the largest $\rho_i(2)$ such that $\alpha \leq 1$. Plugging into the IC constraint provides the final step to determine the optimal mechanism.

Next, we describe the case in which discrimination is absent in the designer's consideration. We start again with a simple corollary to Proposition 2.

Corollary 4. *Suppose $\mathbb{E}[\hat{\Psi}|\mathcal{B}]$ is constant in $\mathcal{B}$. Then a designer, for any ξ, wants to maximize the players' welfare conditional on escalation, $\mathbb{E}[\hat{U}|\mathcal{B}]$.*

Discrimination is absent if the type disadvantage, D_i, is inversely proportional to the mimicked player. Then the expected type profile in $\mathcal{E}$ determines only *which amount* of surplus is distributed (on average). It is immediate from Corollary 4 that also in these cases the maximization of joint surplus and the minimization of escalation coincide.

The following specification describes a (binary) version of the case referred to above. Suppose that $\Theta = \{1,2\}$ and that $U_i(m_i;\theta_i,\mathcal{B}) = \frac{h(\theta_i)}{\rho_i(2)}$ for some strictly increasing function h. In this case, $\mathbb{E}[\hat{\Psi}|\mathcal{B}] = 2\frac{1-p(2)}{p(2)}(h(2)-h(1))$ is constant and the designer maximizes

$$\mathbb{E}[\hat{U}|\mathcal{B}] = \sum_\sigma Pr(\sigma)\sum_i \Big(h(2) + \frac{1-\rho_i(2,\sigma)}{\rho_i(2,\sigma)}h(1)\Big).$$

As above, the designer maximizes the function subject to $\alpha \leq 1$. If constraint $\alpha \leq 1$ is non-binding, the optimal signal σ is degenerate (see Proposition 6 for details).

Outside of the polar cases, the equivalence between escalation minimization and joint surplus maximization does not hold.[18] The discussion below Proposition 2 makes it obvious why. Moreover, our formulation of the problem ($P_{\mathcal{B}}$) provides a simple, yet flexible formulation of the arbitration problem. In particular it illustrates how the properties of the escalation game and the arbitrator's trade-off between discrimination and welfare are connected. In Section 5.1 we connect our formulation to existing results in the literature and argue briefly how our approach provides an intuition for these results as a result of the choice of

[18] See Balzer and Schneider (2019) for a case outside the polar ones in which the two objectives are indeed not equivalent. Zheng and Kamranzadeh (2018) also find that the objectives do not always coincide.

$(\mathcal{A}, u_i, \Theta^2)$.

Finally, our results provide a starting point to the broader design question: Which incentives should we provide to arbitrators and how do they rely on the escalation game? Suppose for example that we start at the first polar case where escalation surplus is a constant $E < 1$. Assume further that measuring settlement rates is easier than evaluating the joint surplus provided. Thus, even if a planner wants to maximize joint surplus she may provide incentives to minimize escalation. If the underlying game changes slightly and E is not a constant any more the two objectives do no longer coincide. The incentives provided are distorted. Proposition 2 and problem $(P_{\mathcal{B}})$ can quantify these distortions.

5 Discussion

In this section we provide examples of escalation games/arbitration problems from the literature (Section 5.1), several extensions and modifications (Section 5.2), and we discuss how to identify the set of binding constraints, the main obstacle to apply our approach (Section 5.3).

5.1 Examples from the Literature

We address the “canonical” claim of the arbitration problem by showing it nests a wide range of models as special cases.

Examples. We provide examples of environments captured by our setting.

Example 1 (Exogenous Cost of Conflict). Our first example covers classic settings such as bilateral trade environments.[19] If settlement fails, parties consume an exogenously given outside option which is type-dependent. Thus, u_i is constant in all arguments but θ_i. Consequently, also U_i is constant in all arguments except θ_i. In turn, the beliefs β_i and $f_{\theta_{-i}}$ are irrelevant. In bilateral trade environments the information channel is shut down entirely. Thus equation $(P_{\mathcal{B}})$ is constant in the information structure. Either full settlement is achievable or no mechanism improves upon a fight.

Example 2 (Conflict as Type-Dependent Lottery). Our second example covers a superset of Example 1. Conflict is a type-dependent lottery, capturing all environments where players' continuation strategies are invariant in the belief $\beta_i(\theta_{-i})$.

[19] In fact the classic setting of Cramton, Gibbons, and Klemperer (1987) with ex-ante equal property rights can be translated to our setting through relabeling. Relabeling follows the approach Compte and Jehiel (2009), section II.B lay out for the bilaterial trade environment

A player's best response depends on her own type only.[20] Examples from the literature in this class are Bester and Wärneryd (2006), Compte and Jehiel (2009), Hörner, Morelli, and Squintani (2015), and Meirowitz et al. (2017).

Formally, type-dependent lotteries imply that $f_{\theta_{-i}}(a_{-i}|\sigma, \mathcal{B}) = f_{\theta_{-i}}(a_{-i})$ and that equilibrium action $a_i^*(\theta_{-i}, m_i, \sigma_i, \mathcal{B})$ is invariant in $(\sigma_i, \mathcal{B})$. Abusing notation we can re-write

$$U_i(\sigma_i, \theta_i, \mathcal{B}) = \sum_{\theta_{-i}} \beta_i(\theta_{-i}|\sigma_i, \mathcal{B}) u_i(\theta_i, \theta_{-i}).$$

The linearity in U_i is particularily helpful under the objective of conflict minimization $\xi = 1$. We show in appendix C that equation $(P_{\mathcal{B}})$ reduces to a linear program due to the linearity of U_i in nests the environments of the above mentioned papers. Applying the belief management approach also provides a different intuition for the existing results.

Example 3 (Private Cost of Conflict). The third example is a superset of Example 1 too. Yet, it extends Example 1 along a different dimension by only assuming that u_i is constant in θ_{-i}. That setting covers all cases in which the ex-post payoff of player i depends on the *action profile* and thus on the opponent's *choices*, but not on the opponent's type. The literature on this type of arbitration is small. In a related paper (Balzer and Schneider, 2019) we analyze optimal arbitration in such an environment.[21] The resale literature (e.g. Gupta and Lebrun, 1999; Calzolari and Pavan, 2006a) considers a very related environment as well (although the model is slightly different).

Formally, let $f(a_{-i}|\sigma, \mathcal{B}) := \sum_{\theta_{-i}} \beta_i(\theta_{-i}|\sigma_i, \mathcal{B}) f_{\theta_{-i}}(a_{-i}|\sigma, \mathcal{B})$. Abusing notation again, we can re-write

$$U_i(\sigma_i, \theta_i, \mathcal{B}) = \max_{a_i \in A_i} \sum_{A_{-i}} u_i(\theta_i, a_i, a_{-i}) f(a_{-i}|\sigma_i, \mathcal{B}).$$

In that class of games, players' actions are determined by their cost functions *and* by the opponent's expected action. As the belief about the opponent's type changes so does the expected distribution over actions. This, in turn, may trigger a chain of events. Player A adjusts her strategy accordingly, player B responds to that, and so on until a new fixed point is found. Deviations are complex in that setting. Best responses depend on information also off the equilibrium path. In particular, a deviating player can gain herself an information advantage.

[20] The literature refers to these as games with ex-post equilibria (Crémer and McLean, 1985).

[21] Zheng and Kamranzadeh (2018) employs considers take-it-or-leave-it bargaining in that envrionment.

Analyzing Example 3 is considerably harder than Example 1 and 2 and the literature on such problems is sparse. Indeed, analyzing these cases using classic techniques requires to solve a mechanism-design problem with a complex information externality.

Using our results, the problem reduces to an information-design problem. The economic trade-off in that information-design problem is immediate. Although solving that problem cannot be avoided, our results provide guidance towards its solution. In light of our results the mechanism is a simple derivative of that solution.

5.2 Extensions and Modifications

We state three modification to our environment and highlight how they influence the description of the problem.

Modification 1 (Veto Leads to Play of a Game). In our model we assume that a veto implies a (possibly type-dependent) exogenous outside option. Here, instead, a veto implies the play of a game $(\mathcal{A}^{\mathcal{V}}, v_i, \Theta^2)$ defined analogously to the game in the escalation event; $(v_a, v_b) : \Theta^2 \times \mathcal{A}^{\mathcal{V}} \to (-\infty, 1]^2$. Likewise, we define $V_i(\theta_i, \mathcal{B}^j)$, where $\mathcal{B}^j$ is the public information structure after a veto by $j \in \{i, -i\}$.[22]

Proposition 3. *If $V_i(\theta_i, \mathcal{B})$ is convex in $\mathcal{B}$, the arbitration problem is isomorphic to the canonical arbitration problem.*[23]

The modification includes parties playing the escalation game $(\mathcal{A}, u, \Theta^2)$ in case of a veto as a special case. That special case is often assumed in the literature (e.g. Schweizer, 1989; Bester and Wärneryd, 2006; Hörner, Morelli, and Squintani, 2015). Our paper Balzer and Schneider (2018) discusses one way to overcome the participation problem even absent the convexity assumption we make in Proposition 3.

Modification 2 (No Transfers). In our model we assume that the arbitrator can impose direct utility transfers between players subject to her budget constraint. Here, instead, we do not allow for direct transfers, that is, $t \equiv 0$.

Without transfers, a reduced-form mechanism (γ, z, Σ) may not be implementable because there is no X that can implement z given γ. A necessary

[22] Private information beyond the distribution of types is irrelevant here.

[23] The convexity assumption on V_i is sufficient, but not necessary. A weaker, yet involved assumption is to concentrate on information structures (p, ρ_{-i}), s.t. ρ_{-i} is the belief player $-i$ holds about i after a veto. With abuse of notation we can replace the convexity assumption by "$V_i(\cdot, (p, \cdot))$ is on the convex closure of V_i's graph with respect to p".

and sufficient condition that a DRM exists that implements the reduced-form DRM follows Border (2007). We define the following terms. For any $Q \subset \Theta^2$ let Q_i:=$\{\theta_i | \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in Q\}$ and $\widetilde{Q}$:=$\{(\theta_A, \theta_B) \in \Theta^2 | \theta_i \notin Q_i$ for i=$\{A, B\}\}$. Moreover, let $P(\mathcal{E}) := \sum_{(\theta_i,\theta_j)} p(\theta_i)p(\theta_j)\gamma(\theta_i, \theta_j)$. We have the following general implementation condition. For all $Q \subseteq \Theta^2$

$$\sum_i \sum_{\theta_i \in Q_i} z_i(\theta_i)p(\theta_i) \leq 1 - Pr(\mathcal{E}) - \sum_{(\theta_A,\theta_B)\in\widetilde{Q}} (1 - \gamma(\theta_A, \theta_B))p(\theta_A)p(\theta_B). \tag{GI}$$

Equations (GI) mirror the general implementation condition from Border (2007).

Proposition 4. *Suppose the mechanism cannot impose direct utility transfers. Then a DRM exists that implements a given reduced-form DRM* (γ, z, Σ) *if and only if the reduced-form DRM satisfies* (GI)*.*

We recommend the following proceeding when facing a problem in which utility transfers are not possible. When constructing the function $M_\Sigma(B)$ include an additional non-negativity constraint for the settlement values, z.

Unfortunately, and common in the literature on reduced-form mechanism design, incorporating the constraints (GI) is tedious. We recommend a guess and verify approach. Ignore the constraints (GI) and compute the optimum. If any constraint in (GI) fails, include it and re-optimize.

In the absence of transfers our model is also isomorphic to a model in which a player's type affects her valuation of the share, x_i.

Proposition 5. *Suppose the value of share x_i is $\varphi(\theta_i)x_i$, $\varphi(\theta_i)$ is increasing, and transfers are not available. The arbitration problem is isomorphic to the canonical arbitration problem.*

If transfers are available and a player's valuation for share x_i is $\varphi(\theta_i)x_i$, then the arbitrator does not need the event $\mathcal{E}$ for screening. Instead, she can screen within $\mathcal{Z}$. The problem collapses to the standard quasi-linear, independent-private-value case (see e.g. Mylovanov and Zapechelnyuk, 2013, Proposition 2).

Modification 3 (Confidential Arbitration). In our model we assume that the arbitrator can send private signals to the players. In particular, she can privately communicate with one party about her communication with the other party. In some settings such communications may be prohibited. Any message a third-party releases to player $-i$ about information previously received from player i has to be available to player i as well.

The following corollary to Proposition 1 describes that case.

Corollary 5 (Public Signals)**.** *Suppose that arbitration is confidential. Then it is without loss of generality that all signals are public, that is in any realization $\sigma_A \equiv \sigma_B$.*

Confidential arbitration reduces the complexity of the signaling space and thus adds structure. It helps us to make statements on the choice, Σ.

Under confidential arbitration we can apply results from the Bayesian Persuasion literature following Kamenica and Gentzkow (2011). Plain concavification (Aumann and Maschler, 1995) does not apply since lotteries may be type-dependent. However, "Lagrange-Concavification" (Doval and Skreta, 2018) does apply. In Appendix B we state the Lagrangian function of the corresponding constrained *maximization* problem. We show that the optimal Σ is obtained from the Lagrangian's concave closure.

Finally, there is an intuitive sufficient condition that determines whether signals are needed at all. Fix the signal to be uninformative, that is, $\Sigma(m_A, m_B) : (m_A, m_B) \mapsto (\emptyset, \emptyset)$. Suppressing the uninformative signal in the notation, we state the following auxiliary problem

$$\min_{B} \frac{\xi + (1-\xi)\,(1-\mathbb{E}[U|B])}{\mathbb{E}[\Psi|B] + \mathbb{E}[U|B] - 1}, \tag{P_B}$$

with B consistent. Problem (P_B) describes problem $(P_{\mathcal{B}})$ prohibiting signals and ignoring $(IC)^A$ and $M(B) \neq 0$.

Proposition 6 (No Signals Needed)**.** *Take a canonical confidential arbitration problem under Assumption 3 and 4. If the solution to (P_B) does not violate any constraint $(IC)^A$ and $M(B) \neq 0$ then it is also the solution to $(P_{\mathcal{B}})$*

Proposition 6 follows because any consistent B can directly be implemented. If no constraint outside the objective binds, the optimal B is on the convex closure of the objective and the Lagrangian. There is no room for signals to exploit the curvature further.

5.3 Identifying Binding Constraints

The formulation in Section 4 provides a problem *conditional on knowing* the binding constraints. The expected utility, U_i, can be a highly non-linear function of the information structure. As a result, it is not easy to identify which constraints bind at the optimum and whether local constraints are sufficient for global constraints.

We provide several results and conditions that help overcoming that tractability issue. We want to emphasize once more that tractability is not a problem of our formulation. To the contrary, our formulation can help to overcome the tractability issues. The difficulty rather stems from the information-design part of the problem. In fact, although information design is conceptually an old question, progress in solving information-design problems on a general level has only been made recently. See, e.g. Mathevet, Perego, and Taneva (2017) and Galperti and Perego (2018) for promising attempts to non-cooperative games.

The difficulty in solving the information-design problem influences the identification of the binding constraints. Thus, at this level of generality we can only provide sufficient conditions. First, we provide a sufficient condition for when local incentive constraints are sufficient. Let $D_i^+(m, \theta, \mathcal{B})$ be the ability disadvantage if $\iota_i(\theta) = \theta + 1$ for all θ.

Proposition 7. *Local upward incentive constraints imply incentive compatibility if the following holds at the unconstrained optimum*

$$\frac{\rho_i(m_i)}{p(m_i)} D_i^+(m_i; \theta_i, \mathcal{B}) \quad \textit{is non-decreasing in } m_i. \tag{4}$$

Even if condition (4) is not satisfied we provide a sufficient condition for local upward incentive compatibility.

Definition 8 (MDR)**.** The game $(\mathcal{A}, u, \Theta^2, \mathcal{B})$ satisfies the monotone difference ratio condition **(MDR)** if $D_i^+(m; \theta_i, \mathcal{B})/D_i^+(m-1; \theta_i, \mathcal{B})$ is non-decreasing in θ_i.

Proposition 8. *Suppose (MDR) holds at the optimum. Local incentive constraints imply (global) upward incentive compatibility.*

Consider the following algorithmic guess and verify approach. First, solve problem $(P_{\mathcal{B}})$ assuming $\iota(\theta) = \theta + 1$. If the solution satisfies (MDR), check if additional downward incentive constraints in $(IC)^A$ are violated. If so, use these constraints to replace one belief in B and solve over the constraint set. Do so until you have found an optimum. Given that (MDR) holds, the algorithm provides a solution.

Independently of whether (MDR) is satisfied, if the optimal solution assuming $\iota(\theta) = \theta + 1$ is monotone in the sense of condition (4), ignored incentive constraints are redundant. While monotone solutions appear intuitive, they cannot be guaranteed for all games. We provide two conditions on the primitives of the escalation game implying monotone solutions. Each condition assumes a *type-separable escalation game.* The payoff function of such a game takes the following form

$$u(a_i; a_{-i}, \theta_i) = \phi(a_i, a_{-i}) - \zeta(\theta_i)c(a_i, a_{-i}),$$

with $\zeta > 0$ decreasing, ϕ, c positive, and strictly increasing in a_i. Moreover, we assume that ϕ is decreasing in a_{-i} and u is concave in a_i. For the sake of the argument, we consider the limiting case of a convex action space. Further we assume that ϕ, c are twice differentiable.[24]

Constant Difference Ratio. Condition (4) holds if the ratio $D^+(m_i; \theta_i, \mathcal{B})/D^+(m_i - 1; \theta_i, \mathcal{B})$ is constant in θ_i. Incentive compatibility is satisfied if and only if the expected escalation probability $\gamma_i(m_i)$ is non-decreasing in m_i, which follows from a monotone hazard rate given a constant difference ratio.

Proposition 9. *The difference ratio is constant if, for given distribution of the opponent's action, the (expected) cost function of a player's best response is separable, that is,* $\mathbb{E}[c(a_i(m_i; \theta_i, \mathcal{B}), a_{-i})|m_i, \mathcal{B}] = h(m_i; \mathcal{B})\tilde{g}(\theta_i)$.

A simple example of such a game is to assume an action space $\mathcal{A} = [0, 1]^2$, $\phi(a_i, a_{-i}) = 1/2 + a_i(1 - a_{-i}) - a_{-i}$, and $c(a_i, a_{-i}) = a_i^2$. Moreover, let $\zeta(\theta) = 1/\theta$. For given distribution of her opponent's action, a player's best response is $a_i(m_i; \theta_i, \mathcal{B}) = (1 - \mathbb{E}[a_{-i}|m_i, \mathcal{B}])\theta_i/2$ which is separable, and so is c.

Non-Constant Difference Ratios. If the difference ratio is non-constant we can specify sufficient conditions. For simplicity we assume that $\xi = 1$ and $\Delta\theta := \zeta(\theta - 1) - \zeta(\theta)$ sufficiently small. The main ingredients to the model to guarantee a monotone solution is that actions are strategic complements. Suppose further that the function ϕ provides a division of the pie, that is, $\phi(a_i, a_{-i}) + \phi(a_{-i}, a_i) = 1$, best responses are continuous, and the hazard rate, $\omega(\theta_i) := \sum_{k=1}^{\theta_i} p(k)/p(\theta_i)$, is non-decreasing. An algorithm close to the one solving Example 2 yields the optimal solution. In in appendix D we sketch that algorithm. A simple parameterization is $\phi(a_i, a_{-i}) = 1/2(1 + a_i - a_{-i})$, and $c(a_i, a_{-i}) = (a_i)^2 + 2Ka_i(1 - a_{-i})$.

[24] Formally, we use our model and assume the distance between any two actions approaches 0. The limit result allows us to exploit envelope arguments and a representation in compact notation.

Appendix

A Proofs

A.1 Proof of Lemma 1

Proof. From the definition of z it follows that (γ, z, Σ) is incentive feasible if and only if (γ, X, t, Σ) is incentive feasible, where (γ, X, t) implies z.

Necessity. The ex-ante expectations of settlement values,

$$\sum_i \sum_{\theta_i} p(\theta_i) \sum_{\theta_{-i}} p(\theta_{-i})(1 - \gamma_i(\theta_i, \theta_{-i}))x_i(\theta_i, \theta_{-i}),$$

cannot exceed the ex-ante probability of settlement

$$1 - \sum_{(\theta_A, \theta_B)} p(\theta_A)p(\theta_B)\gamma(\theta_A, \theta_B).$$

Admissibility implies

$$\sum_i \sum_{\theta_i} p(\theta_i) \sum_{\theta_{-i}} p(\theta_{-i})(1 - \gamma_i(\theta_i, \theta_{-i}))t_i(\theta_i, \theta_{-i}) \leq 0.$$

Together that implies (BB).

Sufficiency. Suppose (z, γ, Σ) is incentive feasible and satisfies (BB). Let $x_A(\theta_i, \theta_{-i}) = 1$ and pick t_A such that

$$z_A(\theta_A) = \sum_{\theta_B} p(\theta_B)(1 - \gamma_A(\theta_A, \theta_B))x_A(\theta_A, \theta_B) + t_A(\theta_A).$$

Further, pick $t_B(\theta_B) = z_B(\theta_B)$. Then,

$$\sum_i \sum_{\theta_i} p(\theta_i) \sum_{\theta_{-i}} p(\theta_{-i})(1 - \gamma_i(\theta_i, \theta_{-i}))t_i(\theta_i, \theta_{-i}) \leq 0$$

and (z, γ, Σ) is admissible. □

A.2 Proof of Proposition 1

Proof. Take an arbitrary arbitration mechanism. At a terminal node of this game outcome $\mathcal{Z}$ or $\mathcal{E}$ realizes and this information is common knowledge among players. Moreover, if outcome $\mathcal{E}$ realized a player uses her private history of play and the public history to draw inference about her opponent's private type and her private history (e.g., higher order beliefs). Then, players use this information to play $\mathcal{E}$. That is, the information structure induced by the play of the arbitration mechanism influences players' utility from outcome $\mathcal{E}$.

From an ex-ante perspective, equilibrium play in the arbitration mechanism implies a type-dependent distribution over outcomes $\mathcal{Z}$ or $\mathcal{E}$ together with a realized private and public history. Each type of a player has access to another types' distribution by imitating her equilibrium strategy. In equilibrium, this is not beneficial. Thus, this distribution can be directly implemented by an incentive feasible DRM. Depending on the report profile, the arbitrator decides which event occurs and sends a potentially random private signal to each player. Players use the signals together with the knowledge about their report, the design of γ, and the function that maps reports into private signals to update the information structure.

Restricting attention to reduced-form mechanisms is without further loss by Lemma 1. Full participation is optimal by Assumption 1. Any on-path veto outcome $V(\theta_A) + V(\theta_B)$ can be replicated inside a reduced-form DRM. □

A.3 Proof of Corollary 1

Proof. Full settlement implies pooling. A full settlement solution is incentive feasible iff z_i constant and weakly larger than $\max_{\theta\in\Theta} V_i(\theta)$. V_i is decreasing and a solution exists iff $V_A(1) + V_B(1) \leq 1$ by (BB). □

A.4 Proof of Lemma 2

Proof. If all beliefs in $\{\beta_A(\cdot|j)\}_{j\in\Theta} \cup \beta_B(\cdot|1)$ have full support the proof is a direct application of Bayes' rule.

Assuming full support, Bayes' rule implies that

$$\beta_i(\theta_{-i}|\theta_i) = \frac{p(\theta_i)p(\theta_{-i})\gamma(\theta_i, \theta_{-i})}{Pr(\mathcal{E})\rho_i(\theta_i)}, \tag{5}$$

with $\rho_i(\theta_i) := Pr(\theta_i|\mathcal{E})$ the likelihood that player i has type θ_i conditional the event $\mathcal{E}$. Consistency implies that

$$\rho_i(\theta_i) = \sum_{\theta_{-i}} \beta_{-i}(\theta_i|\theta_{-i})\rho_{-i}(\theta_{-i}). \tag{6}$$

Now fix a set of probability mass functions $\{\beta_A(\cdot|j)\}_{j\in\Theta} \cup \beta_B(\cdot|1)$ each with full support over the entire type space Θ, but otherwise arbitrary.

Using equation (6) and $\{\beta_A(\theta_B|j)\}_{j\in\Theta}$ we can express any $\rho_B(\theta_B)$ as a (linear) function of the vector $(\rho_A(j))_{j\in\Theta}$. Applying equation (5) to $\beta_B(\theta_A|\theta_B)$ and $\beta_A(\theta_B|\theta_A)$ implies that

$$\rho_B(\theta_B)\beta_B(\theta_A|\theta_B) = \beta_A(\theta_B|\theta_A)\rho_A(\theta_A). \tag{7}$$

Applying equation (7) for $\theta_B = 1$ and any $\theta_A \in \Theta$, and substituting for $\rho_B(1)$ as a function of $(\rho_A(j))_{j\in\Theta}$ determines $(\rho_A(j))_{j\in\Theta}$, and thus $\rho_B(\theta_B)$.

Finally, using equation (5) once more determines the remaining functions $\beta_B(\theta_A|\beta_B \neq 1)$ uniquely.

We want to stress that the proof requires that the initial set of beliefs has full support. Yet, we discuss in the proof of Theorem 1 that restricting ourselves to (limits of) sets with full support is sufficient for the analysis. □

A.5 Proof of Lemma 3

Proof. Take a consistent B. Then, there is a mechanism that implements (B, S) for some S. Thus, there is γ such that B follows from Bayes' rule given γ. Moreover, take any B that follows from Bayes' rule given γ. Then, let $\sigma_i = \emptyset$. The mechanism implements (B, S) with $S = 1$. □

A.6 Proof of Lemma 4

Proof. Given the explanation in the main text directly below Lemma 4, the proof follows from Bergemann and Morris (2016). □

A.7 Proof of Theorem 1

We prove Theorem 1 in steps. Steps 1–4 correspond to Observations 1–4 in the main text. Step 5 proves compactness of the choice set. By Corollary 2, it is sufficient to show that for any Σ there is a one-to-one mapping between the reduced-form mechanism and the "prior" B.

Step 1: Homogeneity. We show that B is homogeneous of degree 0 w.r.t. γ via the following claim.

Claim. γ implements B iff every escalation rule $\hat{g}_{\mathcal{B}} = \alpha\gamma$ implements B where α is a scalar.

Proof. Suppose γ implements B. Homogeneity of Bayes' rule implies that any escalation rule $\hat{g}_B = \alpha\gamma$ implements B. For the reverse suppose $\alpha\gamma$ implements B and set $\alpha = 1$. If γ is an escalation rule it implements B. □

If B is homogeneous of degree 0 w.r.t. γ so is U_i; γ is homogeneous of degree 1 by definition and so is y_i.

Step 2: Most-Costly Escalation Rule. We show that B, for fixed Σ, determines $Pr(\mathcal{E})$. That is, the set of all escalation rules implementing a given information structure, (B, Σ), is defined up to the real numbers $\{\alpha\}$. The escalation probability is linear in any α.

Fix a consistent B and take *some* escalation rule $\hat{\gamma}$ that implements B. Step 1 implies that each escalation rules that implements B satisfies

$$Pr(\mathcal{E}) = \sum_{(\theta_A,\theta_B)} p(\theta_A)p(\theta_B)\alpha\hat{\gamma}(\theta_A, \theta_B).$$

Let $\{\alpha\}$ be the set of all α such that $\forall(\theta_A, \theta_B)$, $\alpha\hat{\gamma}(\theta_A, \theta_B) \leq 1$ and $\hat{\gamma}(\theta_A, \theta_B) = \alpha\hat{\gamma}(\theta_A, \theta_B) \leq 1$. The set $\{\alpha\}$ determines all escalation rules implementing B. Its largest element determines the most-costly escalation rule uniquely.

Step 3a: Set of binding Constraints and Linearity in $\{\alpha\}$. Consider the optimal mechanism.

Claim. For any θ_i the participation constraint or an incentive constraint is satisfied with equality.

Proof. To the contrary, suppose neither the participation constraint nor an incentive constraint holds with equality. Then, we can reduce $z_i(\theta_i)$ until one of the above constraints holds with equality, and all constraints remain satisfied. □

Let Θ^{IC} be the set of types with at least one binding incentive constraint and let Θ^{PC} be the set of types with a binding participation constraint. By the previous claim $\Theta^{IC} \cup \Theta^{PC} = \Theta$. Further, let $\Theta^{I}(\theta_i)$ be the set of types such that θ_i's incentive constraints w.r.t to any $\theta \in \Theta^{I}(\theta_i)$ hold with equality. We say $\widehat{\Theta}_i \subset \Theta_i^{IC}$ describes a *cycle* if for any $\theta_i \in \widehat{\Theta}_i$, it holds that $\theta_i \notin \Theta_i^{PC}$ and $\Theta_i^{I}(\theta_i) \subset \widehat{\Theta}_i$.

Claim. It is without loss of generality to assume no cycles exist.

Proof. Suppose $\widehat{\Theta}_i$ describes a cycle. Reducing $z_i(\theta_i)$ for all $\theta_i \in \widehat{\Theta}_i$ under condition $z_i(\theta_i) - z(\theta_i') = y_i(\theta_i'; \theta_i) - y_i(\theta_i; \theta_i)$ for any $\theta' \in \Theta^I(\theta_i)$ is possible without violating any other constraint since $\Theta_i^I(\theta_i) \cap \{\Theta_i^{PC} \cup \{\Theta_i^I(k)\}_{k \notin \widehat{\Theta}_i}\} = \emptyset$. □

Claim. z_i is linear in α given B.

Proof. Consider $\theta_i \in \Theta_i^{PC}$. Then, $z_i(\theta_i) = V_i(\theta_i) - y_i(\theta_i; \theta_i)$. The first term of the RHS is a constant, the second is linear in α by step 1. For any $\theta_i \in \Theta_i^{IC}$, the incentive constraint is $z_i(\theta_i) = z_i(\theta_i') + y_i(\theta_i'; \theta_i) - y_i(\theta_i; \theta_i)$ if $\theta_i' \in \Theta_i^I(\theta_i)$. Given $z_i(\theta_i')$, linearity holds because y_i is linear in α by step 1. Now, either $\theta_i' \in \Theta_i^{PC}$, or, $z_i(\theta_i')$ is linear given some $z_i(\theta_i'')$ with $\theta_i'' \in \Theta_i^I(\theta_i)$. No cycles exist so that recursively applying the last step yields the desired result. □

Step 3b: Homogeneity of the expected Shares. Using the results from step 3a, let $\mathbb{P}_i(\Theta)$ describe the finest partition of Θ into subsets Θ_i^p such that for every $\theta_i \in \Theta_i^p \in \mathbb{P}_i(\Theta)$ every $\Theta^I(\theta_i) \subseteq \Theta_i^p$. Let $\Theta_i^p(\theta) := \{\Theta_i^p \in \mathbb{P}_i(\Theta) : \theta \in \Theta_i^p\}$ identify the element of the partition to which θ belongs. Finally, let $\hat{\theta}_i(k) := \{\max \theta \in \Theta_i^p(k) : \theta \in \Theta_i^{PC}\}$. By the first two claims of step 3a it is without loss to assume that all objects are well-defined and thus $\hat{\theta}_i$ is non-empty for any $\theta_i \in \Theta$. Using the last claim in Step 3a, we can find a set of functions $H_i(\gamma)$ solving

$$\sum_{\theta_i} p(\theta_i) z_i(\theta_i) = -H_i(\gamma) + \sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i) V_i(\theta_i) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i) V_i(\hat{\theta}_i(\theta_i)). \tag{8}$$

Straightforward algebra implies $H_i(\alpha\gamma) = \alpha H_i(\gamma)$. Thus, $H_i(\alpha\gamma)$ is homogeneous of degree 1 in γ.

Step 4: Determining α via constraint (BB) An arbitration outcome is only admissible if the ex-ante expected settlement values are weakly lower than the probability of settlement, (BB). That is, $\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \leq 1 - Pr(\mathcal{E})$, where the RHS is strictly lower than 1 by Assumption 3. By step 1 any escalation rule $\alpha\gamma$ implements the same B. If each $\alpha\gamma$ is implementable then $\alpha\gamma$ satisfies (BB). By step 3b we can rewrite (BB) as

$$\sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i) V_i(\theta_i) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i) V_i(\hat{\theta}_i(\theta_i)) - 1 \leq \sum_i \alpha H_i(\gamma) - Pr(\mathcal{E}). \tag{BB'}$$

Given Θ_i^{PC}, Θ_i^{IC}, and $\{\Theta_i^I(\theta_i)\}_{\theta_i}$ the LHS is independent of the arbitrator's choice. Moreover, the LHS must be positive at the optimum because $\alpha \to 0$ implies convergence to full settlement which is ruled out by Assumption 3. An immediate result of that is that $\sum_i \alpha H_i(\gamma) > 0$.

If the arbitrator lowers α all constraints continue to hold, but $Pr(\mathcal{E})$ decreases. Redistributing these resources equally among all player types improves upon the proposed mechanism in all three dimensions imposed on the arbitrator's objective by Assumption 2. Thus, such a mechanism cannot be optimal. Equation (BB') and thus equation (BB) hold with equality at the optimal mechanism.

Step 5: Compactness of (B, Σ)**.**

Any signal, s, can be implemented and since the signal space is finite, the set of signals is closed and bounded. For B, the proof of Lemma 2 provides the main argument. The set of full support distributions $\{\beta_A(\cdot|j)\}_{j\in\Theta} \cup \beta_B(\cdot|1)$ is convex, and so is the set of Bs having full support. The next step is to show that the closure of that set can be attained as well. We show this by using the fact that any B on the closure of the set of (full support) Bs can be attained by a sequence of full support Bs , $\{B_n\}_{n\in\mathbb{N}}$. The following two lemmas provide that result and thus compactness of (B, Σ). Moreover, all constraints are weak inequality constraints, such that the set of implementable (B, Σ) is compact.

Lemma 6. *Any B is consistent if and only if it can be approximated by a convergent sequence of consistent B with full support.*

Proof. Take a sequence of consistent $B_n \to B$. B_n is consistent, that is, it implies some function $f : B \to [0,1]^{K\times K}$, such that $f(B_n) = \gamma_n$ with γ_n implementing B_n. Since f is continuous, $\lim_{n\to\infty} f(B_n) = f(\lim_{n\to\infty} B_n) = \gamma$. Consistency implies

$$g_L(B) = g_R(B), \tag{9}$$

where both g_L and g_R are continuous functions $B \to \mathbb{R}$. We can conclude that $g_L(B) - g_R(B) = \lim_{n\to\infty}[g_L(B_n) - g_R(B_n)] = 0$ and B satisfies consistency. This holds because $g_L(B_n) - g_R(B_n) = 0$.

Conversely, take any B that is implemented by some γ. We show that we can find a sequence of interior B that are consistent and converge to B: Let $\hat{\gamma}$ be the escalation rule that implements B. Choose a sequence of escalation rules in the interior that converges to $\hat{\gamma}$. By Bayes' rule every element of the sequence, γ_n, corresponds to some B_n. Moreover, consistency implies that there exists a continuous function, say $f^{-1} : [0,1]^{K\times K} \to [0,1]^{K\times K}$, such that $f^{-1}(\gamma_n) = B_n$ and B_n satisfies consistency. Note that f^{-1} is continuous which implies that $\lim_{n\to\infty} B_n = \lim_{n\to\infty} f^{-1}(\gamma_n) = f^{-1}(\hat{\gamma}) = B$. □

Lemma 7. *Let $\mathcal{O}$ be a continuous function defined on the domain of $\gamma \in [0,1] \cap C$ where C consists of those γ's that satisfy a given set of weak inequality constraints,* *each of which is continuous in* γ*. Then,* $\arg \max\limits_{\gamma \in [0,1] | \gamma \in C} \mathcal{O}(\gamma) = \arg \sup\limits_{\gamma \in (0,1) | \gamma \in C} \mathcal{O}(\gamma)$.

Proof. Without loss of generality suppose the argument that maximizes $\mathcal{O}$, γ^*, gives rise to a non-interior B, B^*. Then, Lemma 6 implies that we can approximate B^* by a convergent sequence of consistent interior Bs. It follows that $\lim_{n\to\infty} \mathcal{O}(B_n) = \mathcal{O}(B^*)$ because $\mathcal{O}$ is continuous in γ (and through Observation 1 continuous in B). Moreover, because the constraints are inequality constraints and continuous in γ (and B), there is n' such that every element B_n with $n > n'$ satisfies the constraints. Therefore, $\max_{\gamma \in [0,1] | \gamma \in C} \mathcal{O}(\gamma) = \sup_{\gamma \in (0,1) | \gamma \in C} \mathcal{O}(\gamma)$ and $B^* = \lim_{n\to\infty} B_n$. Using Lemma 6 we note that for every B_n there is γ_n so that $\lim_{n\to\infty} \gamma_n = \gamma^*$. □

A.8 Proof of Lemma 5

Proof. Consider the resource constraint. Focus on the formulation (BB') in the proof of Theorem 1, step 4. Assume by contradiction that the set of types with binding participation constraint $\Theta_i^{PC} \neq \{K\}$. The LHS of (BB') is $\sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i) V_i(\theta_i) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i) V_i(\hat{\theta}_i(\theta_i)) - 1$. By Assumption 4 this is negative if $\Theta_i^{PC} \neq \{K\}$, contradicting Assumption 3. □

A.9 Proof of Proposition 2

Proof. All references to *steps* refer to those in the proof of Theorem 1. The rule $\overline{g}_B$ is defined in step 2, H_i and $\hat{\theta}(\theta)$ are defined in step 3. Further define $\gamma_i(m_i) := \sum_{\theta_{-i}} p(\theta_{-i}) \gamma_i(m_i, \theta_{-i})$.

At the optimum (BB') holds with equality (step 4). Substituting for $\sum_{\theta_i} p(\theta_i) z_i(\theta_i) = (1 - Pr(\mathcal{E}))$ in problem $(P_{\mathcal{M}})$ implies objective

$$(1-\xi)\Big(1 + \sum_i \sum_\theta p(\theta)\gamma_i(\theta)\widehat{U}_i(\theta;\theta,\mathcal{B})\Big) - \xi Pr(\mathcal{E}). \tag{10}$$

Bayes' rule implies that $p(\theta)\gamma(\theta) = \rho(\theta) Pr(\mathcal{E})$. Factoring out $-Pr(\mathcal{E})$ yields

$$-Pr(\mathcal{E})\Big(\xi - (1-\xi)\sum_i \sum_\theta \rho_i(\theta)\widehat{U}_i(\theta;\theta,\mathcal{B})\Big) = -Pr(\mathcal{E})\left(\xi - (1-\xi)\mathbb{E}[\widehat{U}|\mathcal{B}]\right).$$

For a given B step 1 and 2 imply $\gamma(\theta_A, \theta_B) = \alpha \overline{g}_B(\theta_A, \theta_B) \Rightarrow Pr(\mathcal{E}) = \alpha R(B)$ with $R(B) := \sum\limits_{\theta_A \times \theta_B} p(\theta_A) p(\theta_B) \overline{g}_B(\theta_A, \theta_B)$. (BB') is binding, thus (step 4)

$$\alpha = \frac{Pr(\mathcal{E}) + C}{\sum_i H_i(\gamma)} \tag{11}$$

with constant $C = \sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i)V_i(\theta_i) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i)V_i(\hat{\theta}_i(\theta_i)) - 1$. Substituting for α in $Pr(\mathcal{E}) = \alpha R(B)$ using equation (11) and rearranging implies

$$1 + \frac{C}{Pr(\mathcal{E})} = \frac{\sum_i H_i(\gamma)}{R(B)},$$

and the right hand side is (by step 3 and 4)

$$\sum_i H_i(\gamma)/R(B) = \sum_i \left(\sum_{\theta=1}^{K-1} \rho_i(\iota_i(\theta)) w_i^\iota(\theta_i) D_i^\iota(\iota_i(\theta); \theta, \mathcal{B}) + \sum_{\theta=1}^{K} \rho_i(\theta) \widehat{U}_i(\theta; \theta, \mathcal{B}) \right),$$

which is equivalent to $\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}]$. Thus

$$Pr(\mathcal{E}) = \frac{C}{\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}] - 1}.$$

Substituting into (10) and dividing by C implies

$$\min \frac{\xi - (1-\xi)\mathbb{E}[\widehat{U}|\mathcal{B}]}{\mathbb{E}[\widehat{\Psi}|\mathcal{B}] + \mathbb{E}[\widehat{U}|\mathcal{B}] - 1}.$$

The remaining constraints follow from plugging in for z_i using Theorem 1. Alternatively, we derive them from the Lagrangian in Appendix B. □

A.10 Proof of Proposition 3

Proof. Axiom 3 implies that the identity of the vetoing players becomes common knowledge before players play the veto game.

Fix $\mathcal{M}$ and an equilibrium that implies full participation. Then, $\mathcal{B}$ is relevant only off path. In addition, $\mathcal{B}$ is such that the vetoing player holds his prior beliefs about the non-vetoing player by the don't-signal-what-you-don't-know condition of PBE.

On path vetoes are only relevant if they facilitate participation by others. Participation of i is facilitated if $-i$'s vetoes lower i's expected $V(\theta_i, \mathcal{B})$. By Jensen's inequality that can only happen if V is non-convex in $\mathcal{B}$ (for formal arguments see Celik and Peters, 2011; Balzer and Schneider, 2018, in particular Proposition 2 in the former and Section 3 in the latter). □

A.11 Proof of Proposition 4

Proof. The proof directly follows from Border, 2007, Theorem 3. □

A.12 Proof of Proposition 5

Proof. Assume that players value the share according to $\varphi(\theta_i)$. We show that the model is identical to the main model.

An Auxiliary Model. Assume valuations are reversely ordered, that is, $\theta_1 > \theta_2 > ... > \theta_K$. Define $\check{U}_i(m_i; \theta_i, \mathcal{B}) := U_i(m_i; \theta_i, \mathcal{B})/\varphi(\theta_i)$, where U is the (continuation) payoff from the escalation game. Similarly, transform the outside options $\check{V}_i(\theta_i) := V_i(\theta_i)/\varphi(\theta_i)$ and the values from participating in the mechanism

$$\check{\Pi}_i(m_i; \theta_i) := \Pi(m_i; \theta_i)/\varphi(\theta_i) = z_i(m_i) + \left(\sum_{\theta_{-i}} p(\theta_{-i})\gamma_i(\theta, \theta_{-i})\right) \check{U}_i(m_i; \theta_i, \mathcal{B}).$$

Using the transformed terms only the model is identical to that in the main text. We call it the auxiliary model. What remains to show is that the auxiliary model's solution implies the correct model's solution.

Participation Constraints. If the auxiliary model implies full participation, so does the correct model because

$$\check{\Pi}_i(\theta_i; \theta_i) \geq \check{V}_i(\theta_i) \Leftrightarrow \theta_i \cdot \check{\Pi}_i(\theta_i; \theta_i) \geq V_i(\theta_i) \Leftrightarrow \Pi_i(\theta_i; \theta_i) \geq V_i(\theta_i).$$

Full-settlement in the auxiliary model implies

$$\check{\Pi}_i(1; 1) \geq \check{V}_i(1) \Leftrightarrow z_i(1) \geq \check{V}_i(1) \Leftrightarrow \theta_1 \cdot z_i(1) \geq V_i(1).$$

Resource feasibility requires $\sum_i \check{V}_i(1) \leq 1$. Hence full-settlement in the correct model is admissible if and only if it is admissible in the auxiliary model.

Incentive Constraints. Incentive compatibility in the auxiliary model implies incentive compatibility in the correct model because

$$\check{\Pi}_i(\theta_i; \theta_i) - \check{\Pi}_i(m_i; \theta_i) \geq 0 \ \Leftrightarrow \theta_i \Big(\Pi_i(\theta_i; \theta_i) - \Pi_i(m_i; \theta_i)\Big) \geq 0 \Leftrightarrow \Pi_i(\theta_i; \theta_i) - \Pi_i(m_i; \theta_i) \geq 0.$$

Budget balance holds because all z are identical in both models. □

A.13 Proof of Proposition 6

Proof. We start with a corollary to Lemma 4.

Corollary 6. *Take any distribution over a set of consistent information structures, F($\mathcal{B}$). There exists a consistent information structure $\overline{\mathcal{B}} = (\overline{B}, \overline{\Sigma})$ that* *induces this distribution.*

Applying Corollary 7, confidential arbitration means that the arbitrator can choose any distribution over consistent information structures. Each consistent information structure induces the belief system $B(s)$ which itself is consistent. Following much the same steps from the proof of Proposition 2 the arbitrator's objective becomes $\sum_s Pr(s)\mathcal{O}(s)$, where

$$\mathcal{O}(B(s)) := \frac{\xi + (1-\xi)\left(1 - \mathbb{E}[\widetilde{U}|B(s), s]\right)}{\mathbb{E}[\widetilde{\Psi}|B(s), s] + \mathbb{E}[\widetilde{U}|B(s), s] - 1}, \text{ with}$$

$$\begin{aligned} \rho_i(\theta_i|s) &:= Pr(\theta_i|\mathcal{E}, s), \\ \mathbb{E}[\widetilde{\Psi}|B(s), s] &:= \sum_i \sum_{\theta\in\Theta} \rho_i(\iota_i(\theta_i)|s)\widehat{\Psi}_i(\theta_i, B(s)), \quad \text{and} \\ \mathbb{E}[\widetilde{U}|B(s), s] &:= \sum_i \sum_{\theta\in\Theta} \rho_i(\theta|s)U(\theta; \theta, B(s)). \end{aligned}$$

She minimizes this objective subject to the constraints stated in Problem (P_B) and subject to downward incentive constraints $(IC)^-$. If these additional constraints do not bind, it is easy to see that the optimal signal structure puts full mass on the belief system with the lowest value of $\mathcal{O}$, which proves the proposition.

If an additional constraint binds, the optimal signal can be found by setting up the Lagrangian function of the problem (see Lemma 10 in Appendix B) and choosing the signal distribution that concavifies the inverse of that function. □

A.14 Proof of Proposition 7

Proof. Recall that $\gamma_i(m_i) = Pr(\mathcal{E})\rho_i(m_i)/p(m_i) = \sum_{\theta_{-i}} p(\theta_{-i})\gamma_i(\theta, \theta_{-i})$ is the expected probability of escalation given report m_i. We prove Proposition 7 as a special case of Lemma 8.

Lemma 8. *If $\gamma_i(m)D_i^+(m; k, \mathcal{B})$ is non-decreasing in m on some interval $[\underline{m}, \overline{m}]$ and $k \in [\underline{m}, \overline{m}]$, then local incentive compatibility for type k implies incentive compatibility for any report in that interval.*

Proof. Take k and m. Incentive compatibility holds iff

$$\begin{aligned} & z_i(k) + \gamma_i(k)U_i(k; k, \mathcal{B}) \geq z_i(m) + \gamma_i(m)U_i(m; k, \mathcal{B}) \\ \Leftrightarrow \qquad & -\gamma_i(m)U_i(m; k, \mathcal{B}) \geq z_i(m) - z_i(k) - \gamma_i(k)U_i(k; k, \mathcal{B}). \qquad (12) \end{aligned}$$

Assume first that $m > k$. Adding and subtracting $\sum_{\theta=k+1}^{m} \gamma_i(m)U_i(m;\theta,\mathcal{B})$ to the LHS of (12) turns it into

$$\gamma_i(m)\left(\sum_{\theta=k}^{m-1} D_i^+(m;\theta,\mathcal{B}) - U_i(m;m,\mathcal{B})\right).$$

Adding and subtracting $\sum_{\theta=k+1}^{m-1} z_i(\theta)$ to the RHS of (12) and using local downward incentive compatibility, i.e., $z_i(\theta) - z_i(\theta-1) \leq y_i(\theta-1,\theta-1) - y_i(\theta-1,\theta)$, implies

$$\sum_{\theta=k+1}^{m}(z_i(\theta)-z_i(\theta-1))-\gamma_i(k)U_i(k;k,\mathcal{B}) \leq \sum_{\theta=k}^{m-1}\gamma_i(\theta-1)D_i^+(\theta-1;\theta,\mathcal{B})-\gamma_i(m_i)U_i(m;m,\mathcal{B}).$$

The RHS of the above equation is an upper bound on the RHS of (12). Thus, (12) holds if

$$\sum_{\theta=k}^{m-1}\gamma_i(m)D_i^+(m;\theta,\mathcal{B}) \geq \sum_{\theta=k}^{m-1}\gamma_i(\theta-1)D_i^+(\theta-1;\theta,\mathcal{B}), \tag{13}$$

which holds since $\gamma_i(m)D_i^+(m;\theta,\mathcal{B})$ is non-decreasing in m.

For $m < k$, take equation (12), add and subtract $\sum_{\theta=m+1}^{k-1}\gamma_i(m)U_i(m;\theta,\mathcal{B})$ from the LHS. Iteratively applying local downward incentive compatibility to $z_i(m)$ and simplify to

$$\sum_{\theta=m}^{k-1}\gamma_i(\theta)D_i^+(\theta;\theta,\mathcal{B}) \geq \sum_{\theta=m+1}^{k-1}\gamma_i(m)D_i^+(m;\theta,\mathcal{B}),$$

which again holds since $\gamma_i(m)D_i^+(m;\theta,\mathcal{B})$ is non-decreasing in m. □

The special case of Lemma 8 with $\underline{m} = 1$ and $\overline{m} = K$ for any k concludes the proof of Proposition 7. □

A.15 Proof of Proposition 8

We prove Proposition 8 as a special case of Lemma 9.

Lemma 9. *Local incentive constraints and (MDR) imply that $\gamma_i(m)D_i^+(m,k;\mathcal{B})$ is non-decreasing in m for any $m > k$.*

Proof. Take i and any m and $m-1$. Local incentive compatibility implies

$$y_i(m,m) - y_i(m,m-1) \geq z_i(m-1) - z(m) \geq y_i(m-1,m) - y_i(m-1,m-1).$$

Thus $\gamma_i(m)D_i^+(m;m,\mathcal{B}) \geq \gamma_i(m-1)D_i^+(m-1;m,\mathcal{B})$ or equivalently

$$\frac{\gamma_i(m)}{\gamma_i(m-1)} \geq \frac{D_i^+(m-1;m,\mathcal{B})}{D_i^+(m;m,\mathcal{B})}. \tag{14}$$

The term $\gamma_i(m)D_i^+(m;k,\mathcal{B})$ increases in m if $\gamma_i(m)D_i^+(m;k,\mathcal{B}) \geq \gamma_i(m-1)D_i^+(m-1;k,\mathcal{B})$ or equivalently,

$$\frac{\gamma_i(m)}{\gamma_i(m-1)} \geq \frac{D_i^+(m-1,k;\mathcal{B})}{D_i^+(m,k;\mathcal{B})} \tag{15}$$

which holds by (MDR) and (14) if $m > k$. □

A.16 Proof of Proposition 9

Proof. Assume without loss that $\zeta(\theta_i) = 1/\theta_i$. A player's best-response to her opponent's action, $a(m_i;\theta_i,\mathcal{B})$, satisfies first-order conditions. The envelope theorem implies

$$\begin{aligned} U(m_i;\theta_i,\mathcal{B}) &= U(m_i;1,\mathcal{B}) + \int_1^{\theta_i} c(a_i(m_i;s,\mathcal{B}))/s^2 ds \\ &= U(m_i;1,\mathcal{B}) + h(m_i;\mathcal{B}) \int_1^{\theta_i} g(s)ds, \end{aligned}$$

where $g(s) := \tilde{g}(s)/s^2$ and where we used that $c(a_i(m_i;\theta_i,\mathcal{B})) = h(m_i;\mathcal{B})g(\theta_i)$. Thus, $D_i^+(m_i;\theta_i,\mathcal{B}) = h_i(m_i;\mathcal{B}) \int_{\theta_i}^{\theta_i+1} g(s)ds$. Moreover, for any m_i and m_i' we have that

$$\frac{D_i^+(m_i;\theta_i,\mathcal{B})}{D_i^+(m_i';\theta_i,\mathcal{B})} = \frac{h_i(m_i;\mathcal{B}) \int_{\theta_i}^{\theta_i+1} g(s)ds}{h_i(m_i';\mathcal{B}) \int_{\theta_i}^{\theta_i+1} g(s)ds} = \frac{h_i(m_i;\mathcal{B})}{h_i(m_i';\mathcal{B})},$$

which is independent of θ_i. □

B Lagrangian Problem

Remark. Our argument throughout this section assumes that $\overline{g}_B(K,K) = 1$. This normalization is without loss. For cases in which $0 < \overline{g}_B(K,K) < 1$ relabeling provides the missing step. The remaining cases with $\gamma(K,K) = 0$ are covered by continuity of B in γ. Lemma 7 in the proof of Theorem 1 provides the corresponding formal argument.

The designer's choice is $cs = (z,\gamma)$. The choice set is CS.

Lemma 10. *The Lagrangian approach yields the global optimum.*

Proof. We use Theorem 1 in Luenberger (1969) to show that the Lagrangian approach is sufficient. Let $\mathbf{t}$ be the vector of Lagrangian multiplier. Further, let $G(\cdot)$ be the set of inequality constraints. Define $\mathrm{w}(\mathbf{t}) := inf\{-Obj|cs = (\gamma, z) \in CS, G(cs) \leq \mathbf{t}\}$, where Obj is the objective the designer wants to maximize. The Lagrangian is sufficient for a global optimum if $\mathrm{w}(\mathbf{t})$ is convex.

Assume for a contradiction that $\mathrm{w}(\mathbf{t}_0)$ is not convex at $\mathbf{t}_0$. Then, there is $\mathbf{t}_1$, $\mathbf{t}_2$ and $x \in (0,1)$ such that $x\mathbf{t}_1 + (1-x)\mathbf{t}_2 = \mathbf{t}_0$ and $x\mathrm{w}(\mathbf{t}_1) + (1-x)\mathrm{w}(\mathbf{t}_2) < \mathrm{w}(\mathbf{t}_0)$. For $j \in \{1,2\}$ let $cs_j = (\gamma[j], z[j])$ describe the optimal solution, such that $-Obj(cs_j) = \mathrm{w}(\mathbf{t}_j)$. Note that $\gamma[j]$ induces $\mathcal{B}[j] = (B[j], S[j])$. Then, consider the choice cs_0 such that $z[0] = \lambda z[1] + (1-x)z[2]$, $\gamma[0] = x\gamma[1] + (1-x)\gamma[2]$ and $\Sigma[0] = \{\Sigma[1], \Sigma[2]\}$, with $Pr(\Sigma[1]) = x$. The choice cs_0 corresponds to a belief system $B[0]$ induced by some $\gamma[0]$, Moreover, it is common knowledge that $\Sigma[0]$ induces $\mathcal{B}$ where realization $\mathcal{B}[1]$ occurs with probability x and $\mathcal{B}[2]$ with probability $1-x$. By construction all constraints are satisfied and the solution value equals that of the convex combination

$$\mathrm{w}(\mathbf{t}_0) = -Obj(cs_0) = \sum_{j\in\{1,2\}} Pr(\Sigma[j]) - Obj(\Sigma[j]) = x\mathrm{w}(\mathbf{t}_1) + (1-x)\mathrm{w}(\mathbf{t}_1).$$

A contradiction. □

For any i, θ, the constraints to the minimization problem are

$$\forall\theta \neq \theta' \quad -(z_i(\theta) - z_i(\theta')) - y_i(\theta;\theta) + y_i(\theta';\theta) \leq 0, \tag{IC}$$

$$-z_i(\theta) - y_i(\theta;\theta) + V_i(\theta) \leq 0, \tag{PC_i}$$

$$-1 + \textstyle\sum_i \sum_{\theta=1}^K p(\theta) z_i(\theta) + Pr(\mathcal{E}) \leq 0, \tag{RC}$$

$$\gamma(\theta_A, \theta_B) - 1 \leq 0. \tag{F}$$

We now derive the Lagrangian representation of the optimization problem. First, we state the complementary slackness conditions and the respective Lagrangian multipliers

$$\begin{aligned}
[z_i(\theta) - z_i(\theta') + y_i(\theta;\theta) - y_i(\theta';\theta)]\nu^i_{\theta,\theta'} &= 0, & \nu^i_{\theta,\theta'} &\geq 0;\\
[z_i(\theta) + y_i(\theta;\theta) - V_i(\theta)]\lambda^i_\theta &= 0, & \lambda^i_\theta &\geq 0;\\
\left[1 - \sum_i \sum_\theta p(\theta) z_i(\theta) - Pr(\mathcal{E})\right]\delta &= 0, & \delta &\geq 0;\\
[1 - \gamma(\theta_A, \theta_B)]\mu_{\theta_A,\theta_B} &= 0, & \mu_{\theta_A,\theta_B} &\geq 0.
\end{aligned}$$

For any Lagrangian multiplier, say t, we introduce the following notation $\tilde{t} \equiv \frac{t}{\delta}$.

Define

$$\tilde{\Lambda}^i(\theta) := \sum_{k=1}^{\theta} \tilde{\lambda}^i_k. \tag{16}$$

Next, we characterize the solution in terms of the Lagrangian objective.

Lemma 11. *$\mathcal{B}$ is an optimal solution to the designers problem if and only if there are Lagrangian multipliers that satisfy complementary slackness and $\mathcal{B}$ maximizes*

$$\frac{(1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}]-1}{\widehat{\mathcal{L}}(\mathcal{B})-1},$$

$$\begin{aligned} \textit{where } \widehat{\mathcal{L}}(\mathcal{B}) :=& \mathcal{T}(\mathcal{B}) + \sum_i \Bigg[\sum_{\theta=1}^{K} \rho_i(\theta)\hat{U}_i(\theta;\theta,\mathcal{B}) \\ &+ \sum_{\theta=1}^{K-1} \sum_{\theta'=\theta+1}^{K} \frac{\tilde{\nu}^i_{\theta,\theta'} + \mathsf{M}^{\mathsf{i}}(\theta) - \tilde{\nu}^i(\theta,\theta')}{p(\theta)} \rho_i(\theta)\{\hat{U}_i(\theta;\theta,\mathcal{B}) - \hat{U}_i(\theta;\theta',\mathcal{B})\} \\ &- \sum_{\theta=1}^{K-1} \sum_{\theta'=\theta+1}^{K} \frac{\tilde{\nu}^i_{\theta,\theta'}}{p(\theta')} \rho_i(\theta') \left\{\hat{U}_i(\theta';\theta,\mathcal{B}) - \hat{U}_i(\theta';\theta',\mathcal{B})\right\} \Bigg], \end{aligned} \tag{17}$$

where $\mathsf{M}^{\mathsf{i}}(\theta) := \tilde{\Lambda}^i(\theta) - \sum_{k=1}^{k=\theta} p(k)$*, and* $\tilde{\nu}_i(\theta,\theta') := (\sum_{k=1}^{\theta} \sum_{\tilde{\theta}>\theta}^{K} (\tilde{\nu}^i_{\tilde{\theta},k} - \tilde{\nu}^i_{k,\tilde{\theta}})) - (\tilde{\nu}^i_{\theta',\theta} - \tilde{\nu}^i_{\theta,\theta'})$.

$$\mathcal{T}(\mathcal{B}) := -\sum_{\theta_A \times \theta_B} \frac{\rho_A(\theta_A)\beta_A(\theta_B|\theta_A)}{p(\theta_A)p(\theta_B)} \tilde{\mu}_{\theta_A,\theta_B}. \tag{18}$$

Moreover, the following is true at the optimum:

- *Constraint* (BB) *is always binding, i.e., $\delta > 0$.*
- $\mathsf{M}^i(\theta) = \tilde{\nu}^i(\theta,\theta') + \tilde{\nu}^i_{\theta',\theta} - \tilde{\nu}^i_{\theta,\theta'}$ *for any θ'.*
- *If $\tilde{\Lambda}^i(\theta) - \sum_{j=1}^{j=\theta} p(j) > 0$, then there is at least one type $k \leq \theta$ such that this type's* downward incentive constraint *is* binding. *If in addition the upward incentive constraints are redundant, then $\tilde{\nu}^i_{\theta,\theta'} = 0$ for all $\theta' \geq \theta$.*
- *If $\tilde{\Lambda}^i(\theta) - \sum_{j=1}^{j=\theta} p(j) < 0$, then there is at least one type $k \leq \theta$ such that this type's* upward incentive constraint *is* binding. *If in addition the downward incentive constraints are redundant, then $\tilde{\nu}^i_{\theta,\theta'} = 0$ for all $\theta' < \theta$.*
- *If local incentive constraints are sufficient, then $\tilde{\nu}^i_{\theta,\theta'} = 0$ for any θ such that $\theta' > \theta{+}1$ or $\theta' < \theta{-}1$. Moreover, $\tilde{\nu}^i(\theta,\theta') = \mathsf{M}^i(\theta)$ for any θ, θ' such that $\theta' \neq \{\theta{-}1, \theta{+}1\}$.*

Proof. We manipulate the Lagrangian, $\mathcal{L}$, and derive a more tractable dual problem. We want to minimize $\xi Pr(\mathcal{E})+(1-\xi)\Big(-\sum_i\sum_{\theta=1}^K p(\theta)[z_i(\theta)+\gamma_i(\theta)\hat{U}_i(\theta;\theta,\mathcal{B})]\Big)$. We first relax the problem by replacing $\sum_i\sum_{\theta=1}^K p(\theta)z_i(\theta)$ with $1-Pr(\mathcal{E})$. Factoring out $Pr(\mathcal{E})$ from the objective and applying Bayes rule, the objective becomes $Pr(\mathcal{E})(\xi-(1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}])-(1-\xi)$. Dropping the constant $(1-\xi)$, the Lagrangian reads

$$\begin{aligned}\mathcal{L} = {} & Pr(\mathcal{E})\Big(1-(1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}]\Big)+\delta[-1+\sum_i\sum_{\theta=1}^K p(\theta)z_i(\theta)+Pr(\mathcal{E})] \\ & +\sum_i\sum_{\theta=1}^K[-z_i(\theta)-y_i(\theta;\theta)+V_i(\theta)]\lambda^i_\theta \\ & +\sum_i\sum_{\theta=1}^K\sum_{\theta'\in\Theta\setminus\theta}^K[-z_i(\theta)+z_i(\theta')-y_i(\theta;\theta)+y_i(\theta';\theta)]\nu^i_{\theta,\theta'} \\ & +\sum_{\theta_A\times\theta_B}[\gamma(\theta_A,\theta_B)-1]\mu_{\theta_A,\theta_B}.\end{aligned} \tag{19}$$

Using Theorem 1 and Lemma 2 we optimize over $\{z_i(\cdot),\gamma(K,K),\{\beta_A(\cdot|j)\}_{j\in\Theta}\cup$ $\beta_B(\cdot|1)$, with $\gamma(K,K){:=}Pr(\mathcal{E}|\theta_A{=}K,\theta_B{=}K)$.

Step 1: Eliminating $z_i(\cdot)$ using First-order Conditions. Define $\nu^i_{K+1,K}:=$ $0=:\nu^i_{1,0}=\nu^i_{0,1}$ for ease of notation. The FOC w.r.t. $z_i(\theta)$ are

$$p(\theta)\delta-\lambda^i_\theta+\sum_{\tilde{\theta}\in\Theta_i\setminus\theta}(\nu^i_{\tilde{\theta},\theta}-\nu^i_{\theta,\tilde{\theta}})=0. \tag{20}$$

Summing over all K conditions in (20) and recalling definition (16) yields

$$1=\tilde{\Lambda}^i(K). \tag{21}$$

(20) holds for all θ if and only if

$$\sum_{k=1}^{\theta}\sum_{\tilde{\theta}>\theta}^{K}(\tilde{\nu}^i_{\tilde{\theta},k}-\tilde{\nu}^i_{k,\tilde{\theta}})=-\sum_{k=1}^{\theta}p(k)+\tilde{\Lambda}^i(\theta). \tag{22}$$

Thus, $\sum_{k=1}^{\theta}\sum_{\tilde{\theta}>\theta}^{K}\tilde{\nu}^i_{\tilde{\theta},k}>0$ if $\mathsf{M}^{\mathsf{i}}(\theta)>0$ and vice versa for $\sum_{k=1}^{\theta}\sum_{\tilde{\theta}>\theta}^{K}\tilde{\nu}^i_{k,\tilde{\theta}}$. We solve (20) for $\tilde{\lambda}^i_\theta$ and substitute into (19). We also substitute $\tilde{\nu}^i_{\theta',\theta}=\mathsf{M}^{\mathsf{i}}(\theta)+\tilde{\nu}^i(\theta,\theta')-$ $\tilde{\nu}^i_{\theta',\theta}$ for all $\theta'>\theta$ into (19) and sort terms. Moreover, all terms involving $z_i(\cdot)$ cancel out from (19) via (20).

Step 2: Reformulating the Lagrangian Objective. Given the above necessary conditions, we manipulate the Lagrangian objective to derive a more tractable maximization problem. Using Bayes' rule together with the homogeneity established in the proof of Theorem 1 (step 1), applying algebra and using the first-order conditions it is straightforward to show that (19) admits the following representation

$$\mathcal{L} = Pr(\mathcal{E})(\delta + \xi - (1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}]) - \delta C - \delta \sum_{\sigma} \Pr(\mathcal{E})\widehat{\mathcal{L}}(\mathcal{B}), \tag{23}$$

where C is a constant that is independent of the choice variables and reads

$$C{:=}1 - \sum_i \sum_\theta \tilde{\lambda}^i_\theta V_i(\theta) + \sum_{\theta_A \times \theta_B} \tilde{\mu}_{\theta_A,\theta_B} < 0.$$

Define $\gamma(\theta_A, \theta_B) := Pr(\mathcal{E}|\theta_A, \theta_B)$. From the proof of Theorem 1 (step 1) with $\alpha = \gamma(K, K)$ it follows that $\gamma(\theta_A, \theta_B) = f(\mathcal{B}, \theta_A, \theta_B)\gamma(K, K)$, where $f(\mathcal{B}, \theta_A, \theta_B)$ is a positive real number. Thus, $Pr(\mathcal{E}) = \gamma(K, K)R(\mathcal{B})$ with $R(\mathcal{B}) := \sum_{\theta_A \times \theta_B} p(\theta_A)p(\theta_B)f(\mathcal{B}, \theta_A, \theta_B)$. Plugging into (23) yields

$$\mathcal{L} = \gamma(K, K)R(\mathcal{B})(\delta - (1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}]) - \delta C - \delta\gamma(K, K)R(\mathcal{B})\widehat{\mathcal{L}}(\mathcal{B}). \tag{24}$$

The FOC of (24) w.r.t. $\gamma(K, K)$ is

$$R(\mathcal{B})\Big(\delta + 1 - (1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}] - \delta\widehat{\mathcal{L}}(\mathcal{B})\Big) = 0. \tag{25}$$

By Assumption 3 $R(\mathcal{B}) > 0$ and thus, $\widehat{\mathcal{L}}(\mathcal{B}) - 1 > 0$ if $\gamma(K, K) > 0$. Therefore, $\delta = (1 - (1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}])(\widehat{\mathcal{L}}(\mathcal{B}) - 1)^{-1}$. Substituting into (24) and simplifying yields

$$\mathcal{L} = \frac{(-C)(1 - (1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}])}{\widehat{\mathcal{L}}(\mathcal{B}) - 1}, \tag{26}$$

which is minimized if and only if $\frac{(1-\xi)\mathbb{E}[\hat{U}|\mathcal{B}]-1}{\widehat{\mathcal{L}}(\mathcal{B})-1}$ is maximized.[25]

□

C Example: Type-Dependent Lotteries

In this section we provide an algorithm for the conflict minimization in Example 2. We restrict the environment to be monotone and focus on constant surplus re-

[25] The Lagrangian multipliers are such that C is negative at the optimum. Otherwise (19) and (26) imply that $Pr(\mathcal{E})$ is negative, a contradiction to Assumption 3 or to existence of the following mechanism. Take a degenerate signal distribution and set $\gamma(\theta_A, \theta_B) = 1$ for all type profiles.

duction in case of conflict. That makes ξ irrelevant for our problem. This setting nests the solutions from the literature such as Fey and Ramsay (2011), Hörner, Morelli, and Squintani (2015), and the monotone cases in Bester and Wärneryd (2006). We comment on changes when relaxing these assumptions at the end.

Definition 9 (Lottery)**.** $\mathcal{E}$ is a lottery if $U_i(\theta_i; \theta_i, \mathcal{B}|\theta_{-i})$ is constant in $\mathcal{B}$. The lottery is *0-sum* if $u(\theta_i, \theta_{-i}) + u(\theta_{-i}, \theta_i)$ is constant for all (θ_i, θ_{-i}).

The problem is linear in the joint distribution over type-pairs in the event $\mathcal{E}$, ρ. There is a one-to-one relationship to (consistent) B. Due to the linearity we can ignore signals. We abuse notation and replace $\mathcal{B}$ by the function ρ. To simplify, we assume $V_i(\theta_i) = \sum_i p(\theta_{-i})u(\theta_i, \theta_{-i})$ and impose monotonicity.

Definition 10 (Monotone Lottery)**.** A lottery is monotone if

- $u(\theta_i, \theta_{-i}) - u(\theta_i - 1, \theta_{-i})$ is weakly increasing in θ_i and θ_{-i}, and
- the prior p induces a weakly increasing inverse hazard rate $\omega(\theta_i) := \frac{\sum_{k=1}^{\theta_i} p(k)}{p(\theta_i)}$.

In monotone lotteries $\iota_i(\theta_i) = \theta_i + 1$, i.e., local (upward) incentive constraints imply incentive compatibility. Define

$$\Upsilon(\theta_A, \theta_B) := \omega(\theta_A)\left(u(\theta_A, \theta_B) - u(\theta_A - 1, \theta_B)\right) + \omega(\theta_B)\left(u(\theta_B, \theta_A) - u(\theta_B - 1, \theta_A)\right).$$

Restricting attention (without loss) to the $\xi = 1$ case and plugging into the objective yields

$$2u(K, K) + \sum_{\Theta^2 \setminus (1,1)} \rho(\theta_A, \theta_B)\Upsilon(\theta_A, \theta_B) =: \mathcal{O}(\rho),$$

the objective the arbitrator wishes to maximize. Identifying the highest Υ and setting the corresponding ρ equal to 1 achieves that. If $M(\rho) \neq 0$ for that ρ the problem is solved. Otherwise the optimal solution is not feasible. Instead, the arbitrator has to increase available funds by putting some weight on the second highest Υ as well. For monotone 0-sum lotteries there is a simple algorithm to construct optimal arbitration.

Definition 11 (Top-Down Algorithm)**.** Let Θ^2_+ be the set of type pairs (θ_A, θ_B) such that $\rho(\theta_A, \theta_B) > 0$. Begin by setting $\Theta^2_+ = \emptyset$.

1. Set $\rho(K, K) = 1$ and check if $\rho(K, K) \leq \frac{(p(K))^2(\mathcal{O}(\rho) - 1)}{2V(K) - 1}$. If it holds, terminate. Otherwise continue at 2.
2. Identify the set $\Theta^2_N = \{(\theta_A, \theta_B)|(\theta_A, \theta_B) = \arg\max_{\Theta^2 \setminus \Theta^2_+} \Upsilon(\theta_A, \theta_B)\}$.

(a) Set $\rho(K,K)$ to the solution of

$$\sum_{\substack{(\theta_A,\theta_B)\in\\ \Theta_+^2\cup\Theta_N^2}} \frac{p(\theta_A)p(\theta_B)}{(p(K))^2}\rho(K,K) = 1. \tag{27}$$

(b) Replace $\rho(\theta_A,\theta_B) = \frac{p(\theta_A)p(\theta_B)}{(p(K))^2}\rho(K,K)\ \forall(\theta_A,\theta_B) \in \Theta_+^2 \cup \Theta_N^2$.

(c) Check whether the condition in 1 holds. If it holds, decrease all ρ for the set Θ_N^2 at the expense $\rho(K,K)$ keeping the relation of 2(b) until the condition holds with equality. Then, terminate. If it is violated, repeat step 2.

Proposition 10. *Suppose the escalation game is a monotone 0-sum lottery. Optimal arbitration is the solution to the top-down algorithm.*

Proof. Jointly conditions from step 1 and equation (27) are necessary and sufficient for (BB).[26]

By construction the top-down algorithm point-wise maximizes $\mathcal{O}(\rho)$ subject to (BB). What remains is to show that all ignored constraints are satisfied. We show this using monotonicity, i.e., $\gamma(\theta_i{+}1,\theta_{-i}) \geq \gamma(\theta_i,\theta_{-i}) \Leftrightarrow p(\theta_i)\rho(\theta_i{+}1,\theta_{-i}) \geq p(\theta_i{+}1)\rho(\theta_i,\theta_{-i})$ for all θ_i,θ_{-i}.

Monotonicity trivially holds if $\gamma(K,K) \neq 1$ because it implies $\rho(K,K) = 1$. Thus, assume $\gamma(K,K) = 1$. By Bayes' rule

$$\gamma(\theta_A,\theta_B) = \Big(\rho(\theta_A,\theta_B)\Big)/\Big(p(\theta_A)p(\theta_B)\Big)Pr(\mathcal{E}).$$

When $\gamma(\theta_A,\theta_B) > 0$, then $\gamma(\theta_A{+}1,\theta_B) = 1$ and

$$Pr(\mathcal{E})\rho(\theta_A{+}1,\theta_A) = p(\theta_A{+}1)p(\theta_B), \qquad Pr(\mathcal{E})\rho(\theta_A,\theta_A) \leq p(\theta_A)p(\theta_B).$$

Monotonicity holds since $p(\theta_A)\rho(\theta_A{+}1,\theta_A) \geq p(\theta_A{+}1)\rho(\theta_A,\theta_A)$, and all but the local upward incentive constraints are redundant because

$$\sum_{\theta_{-i}=1}^{K} \left(p(\theta_i')\rho(\theta_i,\theta_{-i}) - p(\theta_i)\rho(\theta_i',\theta_{-i})\right)\left[u(\theta_i,\theta_{-i}) - u(\theta_i',\theta_{-i})\right] \geq 0. \tag{28}$$

Finally, we verify that only the highest type's participation constraint binds at the optimum. It implies that upward local incentive constraints hold with equality.

[26] Substitute $\gamma(\theta_A,\theta_B)p(\theta_A)p(\theta_B)(Pr(\mathcal{E}))^{-1}$ for $\rho(\theta_A,\theta_B)$ in the RHS of (BB) and substitute in the LHS accordingly using the path ι.

We verify the claim by induction. We first show that $\Pi_i(K-1;K-1) \geq V(K-1)$. By local incentive compatibility and $\Pi_i(K;K) \geq V(K)$,[27] we know that $\Pi_i(K-1;K-1) \geq \Pi_i(K,K) - y_i(K;K) + y_i(K;K-1)$. Thus, to show that $\Pi_i(K-1;K-1) - V(K-1) \geq 0$ it suffices to show that $\Pi_i(K,K) - V(K-1) \geq y_i(K;K) - y_i(K;K-1)$. The game is a lottery, and we need

$$\sum_{\theta_{-i}} \left(p(\theta_{-i}) - \gamma_i(K)\beta_i(\theta_{-i}|K)\right)\left(u(K,\theta_{-i}) - u(K-1,\theta_{-i})\right) \geq 0.$$

The last bracket is positive by monotonicity. The first is $(p(\theta_{-i}) - \gamma_i(K)\beta_i(\theta_{-i}|K)) = \left(Pr(K,\theta_{-i}) - Pr(K,\theta_{-i},\mathcal{E})\right)/p(K) \geq 0$.

None of the above argument depends on the specifics of K, thus the induction step to verify $\Pi_i(\theta_i;\theta_i) \geq V(\theta_i)$ for all θ_i follows analogously. □

If we give up the 0-sum element of the lottery, welfare maximization is no longer isomorphic to escalation minimization. Let $W(\theta_A,\theta_B) = u(\theta_A,\theta_B) + u(\theta_A,\theta_B)$ be joint surplus of a type pair (θ_A,θ_B). Define

$$\mathcal{O}'(\rho) := 2u(K,K) + \sum_{\Theta^2 \setminus (1,1)} \rho(\theta_A,\theta_B)(\Upsilon(\theta_A,\theta_B) + W(\theta_A,\theta_B) - W(1,1))$$

The general problem becomes

$$\min_{\rho} \frac{\xi + (1-\xi)\sum_{\Theta^2} \rho(\theta_A,\theta_B)W(\theta_A,\theta_B)}{\mathcal{O}'(p) - 1},$$

s.t. $\rho(K,K) \leq \frac{(p(K))^2(\mathcal{O}'(\rho)-1)}{2V(K)-1}$ and $\sum_{\substack{(\theta_A,\theta_B)\in \\ \Theta^2_+ \cup \Theta^2_N}} \frac{p(\theta_A)p(\theta_B)}{(p(K))^2}\rho(K,K) = 1.$

Without 0-sum and $\xi < 1$ the objective is not linear and thus harder to solve. Yet, it remains that signals are of no help because U remains linear in the information structure leaving no room to exploit the curvature.

D Solution Algorithm for Monotone Mechanisms

Here we provide details behind the results obtained for type-separable escalation games at the end of Section 5. Recall that these escalation games feature the following payoff structure.

$$u(a_i;a_{-i},\theta_i) = \phi(a_i,a_{-i}) - \zeta(\theta_i)c(a_i,a_{-i}). \tag{29}$$

[27] In the optimal mechanism it holds that $\Pi_i(K;K) = V(K)$.

Let a^*_{-i} be the equilibrium action of player $-i$ and define $c(a_i) := \mathbb{E}[c(a_i, a^*_{-i})|m_i, \mathcal{B}]$.

First, we derive the designer's objective if ϕ only distributes the pie without destroying any surplus.

The envelope theorem implies

$$U_i(m_i; \theta_i, \mathcal{B}) = \mathbb{E}[\phi(a_i(m_i; 1, \mathcal{B}), a^*_{-i})|m_i, \mathcal{B}] - C(m_i, \theta_i; \mathcal{B}), \ \textit{where} \tag{30}$$

where $C(m_i; \theta_i, \mathcal{B}) := \int_{\zeta(1)}^{\zeta(\theta_i)} c(m_i; s, \mathcal{B}) d(-\zeta(s)) + \zeta(1)c(m_i; 1, \mathcal{B})$ and let $c(m_i; s, \mathcal{B}) := c(a^*_i(m_i; \theta_i, \mathcal{B}))$ be θ_i's expected cost from his optimal action $a^*_i(m_i; \theta_i, \mathcal{B})$. If $\xi = 1$, optimal arbitration maximizes $\sum_i (\mathbb{E}[U_i|\mathcal{B}] + \mathbb{E}[\Psi_i|\mathcal{B}])$. Since $\phi(a_i, a_{-i}) + \phi(a_{-i}, a_i) = 1$, $\sum_i \mathbb{E}[U_i|\mathcal{B}] = 1 - \sum_i \rho_i(\theta_i)\zeta(\theta_i)c(\theta_i; \theta_i, \mathcal{B})$. Moreover, (30) implies that $D^+_i(m_i; \theta_i - 1, \mathcal{B}) = -C(m_i; \theta_i, \mathcal{B}) + C(m_i; \theta_i - 1, \mathcal{B})$.

As we will show below, in the optimum upward adjacent incentive constraints imply incentive compatibility. Thus, $\iota_i(\theta) = \theta + 1$, $w^\iota_i(\theta) = \sum_{k=1}^{\theta+1} p(k)/p(\theta+1) =: \omega(\theta+1)$, and $D^\iota_i = D^+_i$. Let

$$\tilde{S}(\theta_i, \theta_{-i}; \mathcal{B}) := \sum_i \omega(\theta_i) D^+_i(\theta_i; \theta_i - 1, \mathcal{B}) - \zeta(\theta_i)c(\theta_i; \theta_i, \mathcal{B}).$$

The objective becomes $\mathcal{O}(\rho(\cdot, \cdot)) := \sum \rho(\theta_1, \theta_2)\tilde{S}(\theta_1, \theta_2; \mathcal{B})$. If the type space is sufficiently dense, we can set up an auxiliary problem. We replace $\tilde{S}$ with S being defined as

$$S(\theta_i, \theta_{-i}; \mathcal{B}) := (\omega(\theta_i)\Delta\theta - \zeta(\theta_i))c(\theta_i; \theta_i, \mathcal{B}) + (\omega(\theta_{-i}) - \zeta(\theta_{-i}))c_{-i}(\theta_{-i}; \theta_{-i}, \mathcal{B}).$$

Sufficiency of the Auxiliary Problem. We show that the solution to the auxiliary problem solves the original problem if $\Delta\theta$, for any two adjacent types, is sufficiently small. By the intermediate value theorem we have that

$$\tilde{S}(\theta_i, \theta_{-i}; \mathcal{B}) = \sum_i \omega(\theta_i)\Delta\theta c(\theta_i; \tilde{\theta}_i, \mathcal{B}) - \zeta(\theta_i)c(\theta_i; \theta_i, \mathcal{B})$$

for some $\tilde{\theta}_i \in [\theta_i - 1, \theta_i]$. The objective becomes $\sum_i \tilde{\omega}(\theta_i)c(\theta_i; \tilde{\theta}_i, \mathcal{B}) - \theta_i c(\theta_i; \theta_i, \mathcal{B})$, where $\tilde{\omega} := \omega\Delta\theta$. If $\Delta \to 0$, then $c(\theta_i; \tilde{\theta}, \mathcal{B}) \to c(\theta_i; \theta_i, \mathcal{B})$ and the scores of the auxiliary problem and the original problem coincide. Hence the solutions coincide.
Results. Suppose ω is non-decreasing in θ. Moreover, assume the distance between any two adjacent types, $\Delta\theta := \zeta(\theta - 1) - \zeta(\theta)$, is sufficiently small. We will show that, if the game features strategic complements, upward adjacent incentive constraints are necessary and sufficient for all other constraints. In particular, (4)

is satisfied. Now, we state an algorithm that solves the problem.

We use the general Lagrangian approach from appendix B to develop a solution algorithm for our class of games. We apply it to the the auxiliary problem. We first relax that problem by ignoring all global incentive constraints. Then, the Lagrangian of the reduced-form problem becomes

$$\sum_{\theta_i \times \theta_{-i}} \rho(\theta_i, \theta_{-i}) \Bigg(\left(\sum_{j \in \{i,-i\}} (\tilde{\omega}(\theta_j) - \zeta(\theta_j)) c_j(\theta_j, \theta_j; \mathcal{B}) \right) - \frac{\mu(\theta_i, \theta_{-i})}{p(\theta_i) p(\theta_{-i})} \Bigg), \tag{31}$$

where $\mu(\theta_i, \theta_{-i})$ is the Lagrangian multiplier on the feasibility constraints, i.e.,

$$2V(1)\rho(K,K) \leq (p(K))^2 \left(\mathcal{O}(\rho(\cdot,\cdot)) \right). \tag{32}$$

If that constraint does not bind, then the optimal solution features $\rho(K,K) = 1$. This follows from the complementary nature of the conflict, together with the non-decreasing virtual valuations.

Assume that $\rho(K,K) = 1$ is not feasible, that is, $\mu(K,K) > 0$. Then, the least-constrained solution is not feasible and signals may improve.

We state an algorithm, a top-down version with information revelation, and then argue that this algorithm is optimal.

Algorithm. Define the score of a type profile the following way:

$$\hat{S}(\theta_i, \theta_{-i}) = (\omega(\theta_i) - \zeta(\theta_i)) c_i(\theta_i; \theta_i, \mathcal{B}^{\theta_i, \theta_{-i}}) + (\omega(\theta_{-i}) - \zeta(\theta_{-i})) c_{-i}(\theta_{-i}; \theta_{-i}, \mathcal{B}^{\theta_i, \theta_{-i}}),$$

where $\mathcal{B}^{\theta_i, \theta_{-i}}$ is the belief system that results if each match receives full information. We order type profiles according to their score. If the highest type profile is (θ_i, θ_{-i}), then the next highest type profile is either $(\theta_i - 1, \theta_{-i})$ or $(\theta_i, \theta_{-i} - 1)$ by complements.

Signals might improve because the least-constrained problem is not feasible. Hence given active type profiles, implement the optimal information revelation policy, i.e., that which maximizes the objective given the active type profiles. Check whether the objective satisfies constraint (32). If not, continue to the next highest type profile and repeat the maximization.

Optimality of the Algorithm. The increasing hazard rate and strategic complements imply that higher type profiles have a higher score. Moreover, strategic complements imply that, given the optimal information disclosure policy, type profiles with the highest ex-post scores are the most beneficial ones.

Optimal Information Revelation. To construct the concave hull of the Lagrangian objective, (31), we have to distinguish two cases. Define $c(a_i^*(a_{-i}))$ as that part of the cost that is independent of $-i$'s action. If that function is concave, then we disclose no information. In contrast, if that function is convex, then we disclose full information.

Secondly, observe that the form of the Lagrangian objective (31) implies that given $\rho(\cdot,\cdot)$ the information disclosure that maximizes the least-constrained objective is optimal.

Incentive Constraints. We need to verify that this algorithm satisfies the incentive constraints. That is, $\gamma_i(m_i)D_i^+(m_i;\theta_i,\mathcal{B})$ is weakly increasing in m_i. Both $D_i^+(m_i;\theta_i,\mathcal{B}) = c_i(\theta_i,\theta_i;\mathcal{B})$ and $\gamma_i(\theta_i)$ are weakly increasing in m_i by complementarities and the structure of the optimal solution.

E Information Structures

E.1 Derivation

A player uses all publicly available information to compute her own beliefs about the opponent's potential beliefs and higher-order beliefs. On the equilibrium path these computations are correct. In addition, a player uses her privately obtained information to determine the likelihood that her opponent holds a specific belief (and higher-order beliefs).

Recall that any realization (σ_A,σ_B) provides player i with a private signal $\sigma_i = (s_i, m_i)$. In equilibrium, player i forms a belief about the message σ_{-i} which her opponent receives. That belief itself can be decomposed into two parts. A belief about m_{-i} and one about s_{-i}. We address them in turns.

The belief about m_{-i} coincides with the belief about the payoff type θ_{-i} because $\mathcal{M}$ is incentive compatible. We represent the players' updating procedure as a sequence of two updates. First, the player uses her knowledge of the arbitrator's choice of γ and the observation of event $\mathcal{E}$. Applying Bayes' rule using the prior p leads to a probability mass function $\beta_i(\theta_{-i}|m_i)$. Second, she uses the realization s_i that may contain additional information about θ_{-i}. Applying Bayes' rule again—using $\beta_i(\theta_{-i}|m_i)$ as the prior—leads to a (refined) belief $\beta(\theta_{-i}|m_i, s_i)$.

We proceed with the belief about s_{-i}. Conditional on report profile (m_A, m_B), the belief s_{-i} is entirely determined through the known structure of the signal Σ. We represent it by a cumulative distribution function $S_i(s_{-i}|m_i, s_i, \theta_{-i}) : A_{-i} \to [0,1]$.

We determine the public information structure *before* the realization of the signal, but *after* the event $\mathcal{E}$ is announced. In that case, each player has sent her type report but has not yet obtained σ_i. The player therefore knows her belief about the opponent's payoff type $\beta_i(\theta_{-i}|m_i)$, a joint (conditional) distribution over realizations of her own signal and that type $S_i(s_i, \theta_{-i}|m_i) : A_i \rightarrow [0,1] \times \Theta^2$, and over the opponent's private information $S_i(s_{-i}, \theta_{-i}|m_i, s_i)$. We abuse notation to shorten the description of the latter two to $S_i(\theta_{-i}|m_i) := (S_i(s_{-i}, \theta_{-i}|m_i, s_i), S_i(s_i, \theta_{-i}|m_i))$.

The public information structure contains $(\beta_i(\cdot|\theta_i), S_i(\cdot|\theta_i))$ for all types and players. The set B describes the belief about the state, (θ_A, θ_B), up to second order. The matrix S with element $s_{ij} := S_i(\theta_j|\theta_i)$ describes the belief about (expected) realizations of the private signals up to second order. The public information structure at this point is the tuple $\mathcal{B} := (B, S)$.

E.2 Special Case: Public Signals

A private signal, Σ, may contain a public component. A realization σ_i can contain a public component that reveals more information about the state and the distribution of signals to all players and types simultaneously. For a given information structure, (B, Σ), the public component induces a spread over B. A corollary to Lemma 4 states that any realization of the signal can be directly induced by some mechanism.

Corollary 7. *Take any distribution over a set of consistent information structures, F($\mathcal{B}$). There exists a consistent information structure $\overline{\mathcal{B}} = (\overline{B}, \overline{\Sigma})$ that induces this distribution.*

The special case of public signals implies that Σ contains *only a public component.* An example is confidential arbitration. In that case it is without loss to describe a signal realization by some s, and the players' private signal $\sigma_i = (m_i, s)$. The public information structure after s is $B(s)$. Moreover, Σ describes a mapping from Θ^2 onto a lottery over the set of potential realizations $\{s\}$. Finally, the (expected) common knowledge distribution over types conditional on $\mathcal{E}$ is $\overline{B}$. Each element is $\sum_{k \in \{s\}} Pr(k|\theta_i)\beta_i(\theta_{-i}|\theta_i, k)$. Any $B(s)$ is consistent, and so is $\overline{B}$.

References

Armstrong, M. J. and W. Hurley (2002). "Arbitration using the closest offer principle of arbitrator behavior". *Mathematical Social Sciences* 43, pp. 19–26.

Atakan, A. E. and M. Ekmekci (2014). "Auctions, actions, and the failure of information aggregation". *American Economic Review* 104, pp. 2014–2048.

Aumann, R. J. and M. Maschler (1995). *Repeated games with incomplete information*. Cambridge, MA: MIT press.

Balzer, B. and J. Schneider (2018). "Persuading to Participate: Coordinating on a Standard". *mimeo.*

— (2019). "Managing a Conflict: Optimal Alternative Dispute Resolution". *mimeo.*

Bebchuk, L. A. (1984). "Litigation and settlement under imperfect information". *The RAND Journal of Economics*, pp. 404–415.

Bergemann, D. and S. Morris (2016). "Bayes correlated equilibrium and the comparison of information structures". *Theoretical Economics*, pp. 487–522.

Bester, H. and K. Wärneryd (2006). "Conflict and the Social Contract". *The Scandinavian Journal of Economics* 108, pp. 231–249.

Border, K. C. (2007). "Reduced form auctions revisited". *Economic Theory* 31, pp. 167–181.

Börgers, T. and P. Norman (2009). "A note on budget balance under interim participation constraints: the case of independent types". *Economic Theory* 39, pp. 477–489.

Brown, J. G. and I. Ayres (1994). "Economic rationales for mediation". *Virginia Law Review*, pp. 323–402.

Calzolari, G. and A. Pavan (2006a). "Monopoly with Resale". *The RAND Journal of Economics* 37, pp. 362–375.

— (2006b). "On the optimality of privacy in sequential contracting". *Journal of Economic theory* 130, pp. 168–204.

Carroll, G. D. and I. R. Segal (2018). "Robustly Optimal Auctions with Unknown Resale Opportunities". *mimeo.*

Celik, G. and M. Peters (2011). "Equilibrium rejection of a mechanism". *Games and Economic Behavior* 73, pp. 375–387.

Compte, O. and P. Jehiel (2009). "Veto constraint in mechanism design: inefficiency with correlated types". *American Economic Journal: Microeconomics* 1, pp. 182–206.

Cramton, P., R. Gibbons, and P. Klemperer (1987). "Dissolving a partnership efficiently". *Econometrica*, pp. 615–632.

Crémer, J. and R. P. McLean (1985). “Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent”. *Econometrica* 53, pp. 345–361.

— (1988). “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions”. *Econometrica* 56, pp. 1247–1257.

Doornik, K. (2014). “A rationale for mediation and its optimal use”. *International Review of Law and Economics* 38, pp. 1–10.

Doval, L. and V. Skreta (2018). “Constrained information design: Toolkit”. *arXiv preprint arXiv:1811.03588.*

Dworczak, P. (2017). “Mechanism Design with Aftermarkets: Cutoff Mechanisms”. *mimeo.*

Dworczak, P. and G. Martini (2018). “The simple economics of optimal persuasion”. *mimeo.*

Fey, M. and K. W. Ramsay (2011). “Uncertainty and Incentives in Crisis Bargaining: Game-Free Analysis of International Conflict”. *American Journal of Political Science* 55, pp. 149–169.

Fudenberg, D. and J. Tirole (1988). “Perfect Bayesian and Sequential Equilibria - A clarifying Note”. *mimeo.*

Galperti, S. and J. Perego (2018). *A Dual Perspective on Information Design.* Tech. rep.

Georgiadis, G. and B. Szentes (2018). “Optimal monitoring design”. *mimeo.*

Goeree, J. K. (2003). “Bidding for the future: Signaling in auctions with an aftermarket”. *Journal of Economic Theory* 108, pp. 345–364.

Gupta, M. and B. Lebrun (1999). “First price auctions with resale”. *Economics Letters* 64, pp. 181–185.

Hörner, J., M. Morelli, and F. Squintani (2015). “Mediation and Peace”. *The Review of Economic Studies* 82, pp. 1483–1501.

Jackson, M. O. and M. Morelli (2011). “The reasons for wars: an updated survey”. *The handbook on the political economy of war* 34.

Jehiel, P. and B. Moldovanu (2001). “Efficient design with interdependent valuations”. *Econometrica* 69, pp. 1237–1259.

Kamenica, E. and M. Gentzkow (2011). “Bayesian Persuasion”. *American Economic Review* 101, pp. 2590–2615.

Kolotilin, A. (2018). “Optimal information disclosure: A linear programming approach”. *Theoretical Economics* 13, pp. 607–635.

Kolotilin, A. and A. Zapechelnyuk (2019). “Persuasion meets delegation”. *mimeo.*

Lauermann, S. and G. Virág (2012). “Auctions in Markets: Common Outside Options and the Continuation Value Effect”. *American Economic Journal: Microeconomics* 4, pp. 107–130.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*.

Mathevet, L., J. Perego, and I. Taneva (2017). "On Information Design in Games". *mimeo.*

Meirowitz, A., M. Morelli, K. W. Ramsay, and F. Squintani (2017). "Dispute Resolution Institutions and Strategic Militarization". *Journal of Political Economy*, forthcoming.

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

— (1981). "Optimal Auction Design". *Mathematics of Operations Research* 6, pp. 58–73.

Mylovanov, T. and A. Zapechelnyuk (2013). "Optimal arbitration". *International Economic Review* 54, pp. 769–785.

Olszewski, W. (2011). "A Welfare Analysis of Arbitration". *American Economic Journal: Microeconomics* 3, pp. 174–213.

Pavan, A. and G. Calzolari (2009). "Sequential contracting with multiple principals". *Journal of Economic Theory* 144, pp. 503–531.

Schweizer, U. (1989). "Litigation and settlement under two-sided incomplete information". *The Review of Economic Studies* 56, pp. 163–177.

Spier, K. E. (1994). "Pretrial bargaining and the design of fee-shifting rules". *The RAND Journal of Economics*, pp. 197–214.

Zhang, J. (2014). "Optimal Mechanism Design with Aftermarket Interactions". *mimeo.*

Zheng, C. (2002). "Optimal Auction with Resale". *Econometrica* 70, pp. 2197–2224.

— (2018). "A necessary and sufficient condition for Peace". *mimeo.*

Zheng, C. and A. Kamranzadeh (2018). "The Optimal Peace Proposal When Peace Cannot Be Guaranteed". *mimeo.*