Managing A Conflict*

Benjamin Balzer† Johannes Schneider‡

November 1, 2018

Abstract

We study the design of arbitration aiming to settle conflicts that otherwise escalate to a costly escalation game. Participation in arbitration is voluntary. Players have private information about their strength in the escalation game. The designer fully controls settlement negotiations but does not control the escalation game. We transform the mechanism-design problem of arbitration into the information-design dual problem of belief management. The dual problem identifies how the properties of the escalation game influence the optimal mechanism. We use our general results to study optimal alternative dispute resolution in the shadow of a legal contest.

1 Introduction

Resolving a conflict through an open fight often implies high costs. Thus, it is common to attempt to resolve the conflict before it escalates to a fight. Examples abound. Diplomatic action precedes escalation to a war; alternative dispute resolution precedes litigation; or (mediated) negotiations precede a strike. While generally effective, these resolution attempts seldom guarantee settlement. If resolution fails, information revealed during the attempt becomes strategically relevant within the fighting stage. Parties consider that effect when taking actions during the resolution attempt.

In this paper we address a normative question: Which resolution attempt minimizes the likelihood of a costly fight? We study cases in which parties hold private information about their cost of effort in a fight. A fight is an exogenous escalation game of incomplete information. We take a mechanism-design approach to characterize the optimal resolution process for a broad class of escalation games. We assume that a fight is inevitable absent unanimous participation, and that the mechanism cannot control players' strategic choices

*We thank Eddie Dekel, Philipp Denter, Martin Dumav, Willie Fuchs, Daniel Garrett, Johannes Hoerner, Aditya Kuvalekar, Antoine Loeper, Benny Moldovanu, Volker Nocke, Alessandro Pavan, Martin Peitz, Thomas Troeger, Christoph Wolf, Asher Wolinsky, and audiences in Bonn, Faro, Geneva, Lisbon, Maastricht, Mannheim, Madrid, Melbourne, Milan, Munich, Warwick, and Palma for comments and suggestions. 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

within a fight. Yet, strategies in a fight may depend on information conveyed by the mechanism. The mechanism's cost of eliciting information from players non-trivially interacts with information revelation through the mechanism itself.

We consider the following setting. Two players have private information about their cost of effort in a fight. At the first stage, players decide whether to fight or to engage in arbitration. If both players participate in arbitration, the arbitrator may settle the conflict. If she fails to settle, the conflict escalates to the fight. The arbitrator has full control over the terms of settlement and can directly implement it. Escalation, however, leads to the fight—an exogenously given game of incomplete information. There is only one way to avoid the fight. Both players accept the mechanism *and* the mechanism implements settlement. In case of a fight, players update their beliefs before taking an action. Thus, information obtained before escalation has two effects. First, it influences expectations about the outcome. Second, it influences strategic choices.[1]

Accounting for the interaction between design choices in arbitration and continuation strategies complicates the analysis. The arbitrator simultaneously solves a mechanism-design problem and a (non-trivial) information-design problem. The arbitrator has no direct control over the escalation game. Yet, the arbitrator's choices on the design of the mechanism imply the information structure after escalation. The information structure, in turn, influences the outcome of the escalation game. Standard methods of mechanism design are difficult to apply because continuation payoffs in a fight are non-linear in beliefs.

Results. We have three main results. First, we derive a dual problem to the mechanism-design problem. We call the dual problem belief management. Second, we use belief management to provide a general characterization of the optimal mechanism. Third, we apply our results to legal conflicts and characterize optimal alternative dispute resolution (ADR).

Belief management is an information-design problem with an intuitive structure. Belief management disentangles the value of distributing information from the price of eliciting that information. Thus, it describes the economic trade-off the arbitrator faces. Belief management focuses only on the event of escalation. In contrast, the mechanism-design problem fails to provide that distinction and has to take all stages into account simultaneously. As a consequence, it is harder to solve and its solution is harder to interpret. Thus, the belief-management dual simplifies the analysis. Solving the dual problem is necessary and sufficient to characterize the optimal mechanism.

We characterize the optimal information structure conditional on escalation. It puts equal weight on reducing inefficiencies in the subsequent fight and on discrimination between payoff types. Identifying the two channels—discrimination and efficiency in the escalation game—is the key step to characterize the optimal mechanism. We provide a

[1] We model arbitration, that is, the mechanism satisfies parties' interim participation constraints. Our model nests the arbitration case of e.g. Hörner, Morelli, and Squintani (2015) who focus on the first effect only. Optimal mediation, i.e. including ex-post participation constraints, does not change our results qualitatively which is in line with their findings. For a detailed comparison to the literature on resolution mechanisms see Section 2.5.

sufficient condition when additional signals by the arbitrator are redundant. We also provide a solution algorithm to completely characterize the optimal mechanism under certain sufficient conditions.

As an application we study optimal ADR in legal conflicts. ADR corresponds to the arbitration mechanism. Litigation corresponds to the fight. In litigation disputants provide costly evidence to the court. Whoever provides the most convincing evidence wins the dispute. The joint surplus is reduced by the (joint) investment into evidence provision.

Litigation strategies after a failed ADR attempt are a function of the information obtained in ADR. Existing models in the literature on conflict resolution assume that optimal behavior after escalation is independent of the information structure. In contrast, strategies in litigation are sensitive to the information structure. This difference changes results qualitatively.

Optimal ADR implies that (i) the information structure upon escalation is asymmetric, (ii) disputants cannot influence what they learn from the escalation decision, and (iii) any type profile escalates with positive probability. In contrast, optimal arbitration in the literature implies that the information structure is symmetric, disputants can influence what they learn, and some type profiles settle with certainty. We address each result separately.

ad (i): First, asymmetry reduces inefficiencies in litigation. The more symmetric the information structure, the more disputants invest into evidence provision. Asymmetries reduce aggregate investment and thereby increase expected welfare. Higher welfare in litigation reduces the cost of failed ADR. That reduction facilitates initial agreement to ADR. At the optimum, litigation after failed ADR is less inefficient than litigation without an ADR option. In reality, disputants indeed report that ADR is helpful even if it turns out to be unsuccessful (Anderson and Pi, 2004; Genn, 1998).

ad (ii): Second, not being able to influence their learning prevents disputants from using ADR only to extract information. That property eases disputants' incentive constraints. On the equilibrium path, escalation is informative. Disputants know the probability with which each type profile escalates. Suppose that a weak disputant A mimics a strong type during ADR. Upon escalation, she holds the on-path belief of a strong type. If that belief differs from a weak type's on-path belief, it prompts a different continuation strategy, too.

The opponent, B, and the arbitrator cannot detect A's deviation. B continues 'as if' on path. Her strategy is not a best response to A's continuation strategy. The reason is that higher-order beliefs are not common anymore. A is aware of B's belief, and thus of B's strategy. B, in turn, is unaware that A has deviated. She neither expects A to hold an off-path belief nor to follow an off-path strategy.

The lack of common knowledge gives the mimicker an information advantage. She best responds holding correct higher-order beliefs. In contrast, the non-deviator incorrectly assumes on-path escalation and best responds holding incorrect higher-order beliefs. If disputants cannot influence what they learn in ADR, the information advantage vanishes. If beliefs are the same on-path and off-path, continuation strategies are the same, too.

ad (iii): Third, the property that any type profile can escalate ensures no learning. By contrast, suppose the arbitrator always settles two weak types. Then weak types have large incentives to mimic strong types. Upon escalation weak types that mimic strong types are perceived as strong. But then, by the logic from the previous paragraph, weak types obtain an information advantage from mimicking. An information advantage is never optimal.

Related Literature. We contribute to the literature on conflict-resolution mechanisms. Existing models (e.g. Bester and Wärneryd, 2006; Hörner, Morelli, and Squintani, 2015; Spier, 1994) abstract from strategic effects of information revelation. Our model nests these models as a special case. Zheng (2018) uses the all-pay auction as escalation game and provides necessary and sufficient conditions when settlement can be guaranteed, a case we refer to as first best. We state a condition in his spirit. However, we are mainly interested in the complementary case where first best cannot be achieved. Different to the first-best case, effects on behavior require non-standard techniques to characterize the second-best mechanism. We complement Meirowitz et al. (2017). They study interaction of resolution mechanisms with prior behavior, we look at the interaction with subsequent behavior. We compare our results to approaches in the literature in Section 2.5.

Our model includes interdependent valuations and information externalities similar to Jehiel and Moldovanu (2001). Our condition on the existence of first-best mechanisms is in line with Compte and Jehiel (2009). We add an effect on behavior to these models. That addition implies that payoffs become non-linear in beliefs. We show that the departure from linearity implies different results and requires a different analysis.

The models on common agency by Calzolari and Pavan (2006a,b) and Pavan and Calzolari (2009) introduce an effect of design choices within a mechanism on action choices outside that mechanism. More recent, the literature on aftermarkets (Atakan and Ekmekci, 2014; Dworczak, 2017; Lauermann and Virág, 2012; Zhang, 2014) revisits that effect. While related on a higher level, our approach differs. We emphasize how *the distribution and relevance of skills* in the escalation game drives the behavior within and the design of the optimal mechanism. In addition, our main goal is to identify the optimal mechanism if the same parties meet *within* the mechanism *and* the escalation game.

Our *belief-management approach* to the mechanism-design problem is related to the approach of Grossman and Hart (1983) to contracts. Similar to them, we aim to separate benefits from costs. In a first step, we determine the designer's cost of implementing a particular information structure. In the second step, we determine a program to select the least costly information structure. The second-step program shares strong similarities to the formulation of the optimal auction problem in Myerson (1981). Our results allow us to separate costs and benefits of manipulating the information structure.

The arbitration problem is a mechanism-design problem with adverse selection and moral hazard a la Myerson (1982). Our belief-management representation transforms it into a (pure) information-design problem a la Bergemann and Morris (2016). The belief-based techniques of Mathevet, Perego, and Taneva (2017) directly apply to the reformulated problem. Our analysis emphasizes that information-design techniques are of first-order importance when mechanisms have endogenous outside options.

Roadmap. We organize the paper as follows. In the first part, Section 2, we study a stylized model of legal disputes. In the second part of the paper, Section 3 and 4, we derive a characterization of the optimal mechanism for a broad class of escalation games. In Section 5 we discuss various extensions. Section 6 conludes. Proofs are in the appendix.

2 A Model of Alternative Dispute Resolution

In this section we discuss a stylized model of Alternative Dispute Resolution (ADR) in the shadow of formal litigation. The stylized model is illustrative and delivers most of the intuition behind the general model. Allowing for changes in behavior upon escalation implies some complexity already in our stylized model. For the sake of brevity, we omit a discussion on the institutional background of ADR and on technical subtleties. Appendix C provides these discussions and the formal proofs of our statements. Our characterization of optimal ADR is a direct consequence of the general results which we present in Section 3 and 4.

We use the canonical model of resolution mechanisms (e.g. Bester and Wärneryd 2006; Hörner, Morelli, and Squintani 2015). We alter, however, the properties of the escalation game. We model litigation as a legal contest in which disputants provide costly evidence to a judge or jury. Optimal strategies in contests are sensitive to beliefs. Thus, any information disputants obtain during ADR influences their strategies in litigation after a failed ADR attempt.

We structure this section as follows. We first present and analyze our stylized model. Then, we discuss results and relate them to previous findings in the literature.

2.1 Stylized Model

Setting. Two ex-ante identical disputants, A and B, are in a legal dispute. The value of winning the dispute is normalized to 1. The default way to solve the conflict is through formal litigation. However, disputants can participate in a given ADR mechanism that aims for a settlement solution instead. Disputants avoid litigation only if both participate in ADR *and* ADR settles the conflict. If at least one disputant vetoes ADR, her veto decision becomes public and disputants engage in litigation. Similarly, if ADR fails to settle the conflict, failure becomes public, and the conflict escalates to litigation.

Litigation. Litigation is a legal contest with private information. Disputants compete in providing evidence to a judge or jury. Disputant i chooses the quality level of the evidence she provides, $a_i \in [0, \infty)$. The highest quality of evidence wins the lawsuit.

Disputants are privately informed about their marginal cost of increasing the quality of evidence. We denote a disputants private information by her type θ_i. In our stylized model, we focus on the binary case. A disputant is either "strong" ($\theta_i = 1$) or "weak"($\theta_i = K > 2$). Types are independent draws from the same distribution, where p is the probability that $\theta_i = 1$. Non-trivial ADR requires that $p \in (0, \overline{p})$ with $\overline{p} := (K-2)/(2(K-1))$.[2] We refer to the realization (θ_A, θ_B) as a *match* of θ_A and θ_B.

Type θ_i's ex-post utility from evidence profile (a_i, a_{-i}) is

$$u(a_i; a_{-i}, \theta_i) = \begin{cases} 1 - \theta_i a_i & \text{if } a_i > a_{-i} \\ -\theta_i a_i & \text{if } a_i < a_{-i} \\ 1/2 - \theta_i a_i & \text{if } a_i = a_{-i}. \end{cases}$$

ADR mechanisms. ADR is a mechanism offered by an ex-ante uninformed, non-strategic third party—the arbitrator—at the beginning of the game. Because the arbitrator cannot control choices in litigation, a general version of the revelation principle (Myerson, 1982) applies to our problem. It is sufficient to restrict attention to direct-revelation mechanisms with full participation.[3] This implies in particular that any veto is an off-path event.[4]

The arbitrator has limited control over the environment and a mechanism results in either settlement or escalation. Settlement directly implements a sharing rule $x_i(\theta_i, \theta_{-i}) \in [0, 1]$. It determines the settlement allocation if ADR succeeds in avoiding litigation. The arbitrator can destroy parts of the surplus but cannot induce additional welfare. Thus, the arbitrator's budget constraint is $x_A(\theta_A, \theta_B) + x_B(\theta_B, \theta_A) \leq 1$.

The arbitrator can induce litigation by failing to settle but cannot control disputants behavior in subsequent litigation. That property of the mechanism is captured by the escalation rule $\gamma_i(\theta_i, \theta_{-i}) \equiv \gamma_{-i}(\theta_{-i}, \theta_i) \in [0, 1]$. It determines the likelihood that the conflict escalates to litigation. In addition, the arbitrator can release informative signals to disputants. In our stylized model, these signals are often superfluous. We omit a formal description here for the sake of simplicity. We discuss additional signals when presenting the optimal mechanism. Without additional signals, a direct-revelation mechanism $\mathcal{M}$ is

[2] If $p > \overline{p}$, ADR can guarantee settlement by offering an equal split of the pie.

[3] Full participation is optimal because payoffs in litigation under the prior are on the convex closure of the payoff function w.r.t. information structures. See Celik and Peters (2011) for details.

[4] It may appear that a simpler mechanism solves our problem: the arbitrator proposes a single sharing rule that disputants accept or veto. Litigation follows iff either disputant vetoes. In such a setting, the beliefs following a veto depend on the "mechanism's" design (i.e., the proposed sharing rule). While appealing by its simplicity, it is a priori not clear if a simple bargain can implement the optimal mechanism. We apply the revelation principle to obtain the optimal mechanism. It turns out that a *simple bargain cannot implement the optimal mechanism.*

The design approach is helpful. Invoking the revelation principle disentangles ADR design from veto beliefs. Full participation is optimal and a veto triggers an off-path node. (Off path) beliefs are restricted by the equilibrium concept (PBE) only. Veto beliefs on the deviator are type-independent but otherwise arbitrary, those on a non-deviator coincide with the prior. Both restrictions follow from 'no-signaling-what-you-don't-know.'

In general, *one should not confuse off-path veto-beliefs in ADR with on-path veto-beliefs in a simple bargain.* We compare the results of these different approaches in detail in Section 2.5.

a mapping $\Theta^2 \to [0,1] \times [0,1]^2$.

Timing. First, disputants privately observe their types and the ADR mechanism is publicly announced. Then, disputants simultaneously decide whether to participate in ADR. By participating, they commit to accept the outcome of ADR. Upon mutual participation, disputants send a message, m_i, to ADR and ADR either implements settlement or the conflict escalates. Settlement ends the game and disputants enjoy their settlement shares x_i. If the conflict escalates, disputants update their beliefs and decide on their strategy in litigation.

Information Structure. We denote the public information structure conditional on the outcome 'escalation' by $\mathcal{B}$. It is a collection of individual beliefs $b_i(\theta_i)$. An individual belief $b_i(\theta_i) := Pr(\theta_{-i} = 1 | m_i = \theta_i)$ determines the on-the-equilibrium-path likelihood that disputant i's type θ_i attributes to disputant $-i$ being strong.

The escalation rule γ_i implicitly determines that information structure. The outcome 'escalation' is common knowledge and informative. Its likelihood depends on disputants' reports in previous arbitration. Disputants use the implied information to update their beliefs. Since γ_i depends on type reports, different types may have different beliefs about the opponent's type distribution. Thus, $b_i(\theta_i)$ may vary in θ_i. On the equilibrium path the collection $\mathcal{B} = \bigcup_{i,\theta_i} b_i(\theta_i)$ captures all information conveyed through the outcome 'escalation.'

Objective and Solution Concept. We jointly look for an ADR mechanism and an associated equilibrium of the grand game to minimizes the likelihood of litigation. Our solution concept is perfect Bayesian equilibrium (Fudenberg and Tirole, 1988). To fix ideas, assume passive off-path beliefs.[5]

To solve for optimal ADR, we have to first characterize the continuation game of litigation. Thereafter we characterize the optimal mechanism.

2.2 The Continuation Game after Escalation

Litigation is the continuation game in case of escalation. Litigation can occur both on the equilibrium path and off the equilibrium path. Disputants best respond to their information set and expect their opponent to do the same.

In our description we focus on the case in which litigation occurs because ADR resulted in escalation. The case of vetoes follows analogously. We come back to it when we describe the optimal mechanism.

Independent of whether escalation happens on path or off path, each disputant's *information set* consists of two parts. The public part $\mathcal{B} = \{b_A(1), b_B(1), b_A(K), b_B(K)\}$ and a private part (m_i, θ_i). While the public part follows from the properties of the ADR

[5]Choosing passive off-path beliefs is optimal in our setting. Any other off-path belief satisfying the intuitive criterion leads to an equivalent result. In addition, it will become clear as we move on that off-path beliefs are only relevant if some disputant vetoes. The concerns discussed in Sugaya and Wolitzky (2017) are therefore not relevant for our problem.

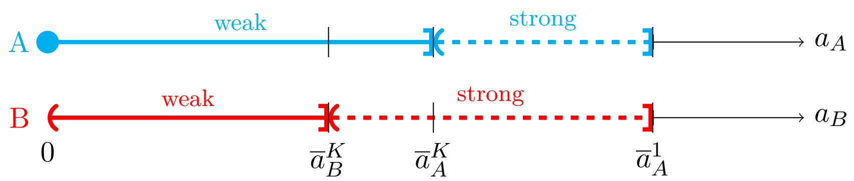


Figure 1: Equilibrium quality of evidence in the on-path continuation game. *All types (piece-wise uniformly) mix. Solid lines denote the intervals of θ_i=K, dashed lines those of θ_i=1. The equilibrium distribution is non-atomistic, apart from the mass point at 0 for disputant A, type K.*

mechanism, the private part consists of the exogenously given information θ_i and the report sent to the mechanism before, m_i.

Suppose disputant i reported $m_i = 1$. In any continuation game, her best response depends on her belief about $-i$'s type distribution $b_i(m_i)$, and her expectations about the associated best response by each type. Thus, also $-i$'s beliefs, $b_{-i}(1)$ and $b_{-i}(K)$ matter for i's decision. Moreover, $-i$ cares about $b_i(\theta_i)$ for the same reasons. That, in turn, implies that both $b_i(1)$ and $b_i(K)$ are relevant for any m_i. As $\mathcal{B}$ is common knowledge, it captures all higher-order beliefs.

Given $\mathcal{B}$, a type's equilibrium strategy in the continuation game is a cumulative distribution function $F_i^{\theta_i}$ over actions. Ignoring ties, the associated expected continuation utility is[6]

$$U_i(m_i; \theta_i, \mathcal{B}) = \sup_{a_i} \Big(F_{-i}(a_i|m_i) - \theta_i a_i \Big), \tag{1}$$

with

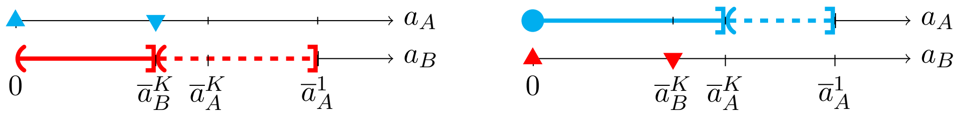
$$F_{-i}(a_i|m_i) = b_i(m_i) F^1_{-i}(a_i) + (1 - b_i(m_i)) F^K_{-i}(a_i). \tag{2}$$

On the Equilibrium Path. On the equilibrium path disputants behave truthfully and $m_i = \theta_i$. To build intuition we make an informed guess. We focus on information structures that have "full support" and imply a monotonic equilibrium. That is, any match occurs with positive probability and strong disputants expect to win more often than weak ones.

The all-pay winner-takes-it-all structure of litigation implies ex-post regrets for any outcome. One consequence of that property is that the equilibrium is in mixed strategies. Strong disputants choose higher quality levels than weak disputants and all disputants pick their equilibrium quality levels from a connected set. No disputant wants to provide evidence strictly above the other disputant's maximum quality level. Therefore, both disputants agree on an upper bound on quality. For a similar reason the equilibrium strategies' support has no holes. We sketch it in Figure 1.

Lemma 1. *Assume $1 > b_A(1) \geq b_B(1) > 0$. A monotonic equilibrium is characterized by quality levels $\overline{a}_A^1 > \overline{a}_A^K \geq \overline{a}_B^K > 0$ that partition the action space. Disputants uniformly mix within a partition. The support of each disputant's strategy is on the intervals*

[6]Ties are relevant neither on nor off the equilibrium path. On the equilibrium path, the supremum coincides with the maximum. However, off path the maximum may not exist.



(a) Quality of evidence in the continuation game after disputant A, type K deviates. *The deviator chooses action ▲ if $b_A(1) > b_A(K)$ and ▼ if $b_A(1) < b_A(K)$. The non-deviator follows her equilibrium strategy.*

(b) Quality of evidence in the continuation game after disputant B, type K deviates. *The deviator chooses action ▲ if $b_B(1) > b_B(K)$ and ▼ if $b_B(1) < b_B(K)$. The non-deviator follows her equilibrium strategy.*

Figure 2: Continuation strategies for different histories if $b_A(1) \geq b_B(1) \neq b_B(K)$.

- $(0, a_A^K]$ *for disputant A, type* K*, and* $(a_A^K, a_A^1]$ *for disputant A, type 1,*
- $(0, a_B^K]$ *for disputant B, type* K*, and* $(a_B^K, a_A^1]$ *for disputant B, type 1.*

In addition, disputant A's type K *has a mass point at* 0 *if* $b_A(1) > b_B(1)$*.*

Further details and closed-form expressions of equilibrium objects are in appendix C. A detailed construction and discussion of the equilibrium is in Siegel (2014).

Off the Equilibrium Path. Off-path escalation follows if a disputant misreports her type during ADR and the outcome is escalation. Mimicking alters beliefs. For example, if type K reports $m_i = 1$, she enters escalation with belief $b_i(1)$ instead of $b_i(K)$. Fix the non-deviating disputant $-i$'s type-specific distribution of actions at $F_{-i}^{\theta_{-i}}$. Nonetheless, changing m_i changes i's winning probability, $F_{-i}(a_i|m_i)$, through $b_i(m_i)$. Through equation (1) that may cause a change in the deviator's optimal continuation strategy.

Such deviations are undetected and the public information structure remains at $\mathcal{B}$. That implies an *information advantage* to the deviator in litigation. Disputant $-i$ does not respond to i's changes in a_i. She keeps her equilibrium distribution of actions $F_{-i}^{\theta_{-i}}$. Conceptually, the deviator—aware of the deviation—best responds to correct expectations. The non-deviating opponent, to the contrary, responds assuming on-path escalation. If the deviator alters her strategy, the non-deviator best responds to incorrect expectations. We graph optimal strategies in the off-path game in Figure 2(a) and 2(b). An information advantage in litigation is beneficial because optimal strategies are information sensitive.

Lemma 2. *Suppose that* $b_i(1) \neq b_i(K)$*. A deviator's optimal action in the continuation game is a singleton and there is a type* θ_i *such that* $U_i(m_i \neq \theta_i; \theta_i, \mathcal{B}) > U_i(\theta_i; \theta_i, \mathcal{B})$*.*

2.3 Optimal ADR

We want to minimize escalation. No full-settlement mechanism exists if $p \in (0, \overline{p})$. Instead any incentive-compatible mechanism leads to escalation with positive probability, and the arbitrator's control ceases once the conflict escalates.[7]

Suppose θ_i reports m_i during ADR. Escalation occurs with probability $\gamma_i(m_i) := p\gamma_i(m_i, 1) + (1-p)\gamma_i(m_i, K)$. The disputants (continuation) value from participation is

[7] ADR *cannot* implement the allocation in the litigation game directly. It cannot control disputants' actions. However, the arbitrator can release an additional signal to influence $\mathcal{B}$ and thereby continuation strategies. To keep the exposition simple, we postpone the discussion of such signals.

$$\Pi_i(m_i;\theta_i) = \underbrace{p\gamma_i(m_i,1)x_i(m_i,1) + (1-p)\gamma_i(m_i,K)x_i(m_i,K)}_{=:z_i(m_i)\text{ (settlement value)}} + \underbrace{\gamma_i(m_i)U_i(m_i;\theta_i,\mathcal{B})}_{=:y_i(m_i;\theta_i)\text{ (escalation value)}}.$$

Alternatively, she can veto the mechanism. In our construction, a veto is an off-path event. By the 'no-signaling-what-you-don't-know' condition of perfect Bayesian equilibrium, the vetoing disputant cannot learn from her own veto decision and thus keeps the prior belief. Non-deviators keep the prior belief due to the assumption of passive beliefs. Through Lemma 1 we can calculate V_i, the value of vetoing. It corresponds to the expected utility in litigation under priors,

$$V_i(\theta_i) = \begin{cases} (1-p)(K-1)/K & \text{if } \theta_i = 1 \\ 0 & \text{if } \theta_i = K. \end{cases}$$

Let $Pr(\mathcal{E})$ be the probability of escalation. Optimal ADR solves

$$\begin{aligned} &\min_{(\gamma_i,x_i)} Pr(\mathcal{E}) \quad \text{s.t. } \forall \theta_i, i, m_i: \\ &\Pi_i(\theta_i;\theta_i) \geq \Pi_i(m_i;\theta_i), \\ &\Pi_i(\theta_i;\theta_i) \geq V_i(\theta_i). \end{aligned} \tag{P$_{min}$}$$

The first set of constraints are incentive constraints. They are binding for weak types as they want to appear strong. The second set of constraints are participation constraints. They are binding for the strong types as they need to be incentivized to participate.

The escalation rule, γ_i, serves a dual purpose in this problem. On the one hand it determines the likelihood that escalation occurs on the equilibrium path. On the other hand it pins down the information structure in the litigation game following escalation. Both aspects are important from the arbitrator's perspective. The first aspect directly affects the performance of the mechanism as the arbitrator wants to minimize escalation. The effect of the second aspect is more subtle. By influencing the information structure $\mathcal{B}$, γ_i affects disputants' expected utility from litigation after escalation. These utilities are key to satisfy both the participation and the incentive constraints.

Conceptually, γ_i serves both a mechanism-design purpose and an information-design purpose. The dual purpose complicates the analysis. The information-design purpose of γ_i is through the implied information structure $\mathcal{B}$. The two objects are connected via Bayes' rule. The information structure, in turn, affects $U_i(m_i;\theta_i,\mathcal{B})$ in two ways.

First, the type distribution in litigation matters. Equation (1) captures this channel. The individual belief $b_i(m_i)$ determines the distribution disputant i faces in litigation when reporting m_i to the arbitrator. That channel is linear in $\mathcal{B}$ and standard in the literature on conflict resolution. It has a direct influence on U_i.

Second, strategic considerations of deviators matter. Equation (2) captures this channel. Beliefs determine continuation strategies and—through equilibrium—best responses to them. An optimal action $a_i(\mathcal{B})$ and the associated distribution $F_{-i}^{\theta_{-i}}(a_i)$ depend on the entire information structure $\mathcal{B}$. That channel is typically non-linear and ignored in most of the literature on conflict resolution. It has an indirect influence on U_i via (a_i, a_{-i}).

From a technical perspective that second channel complicates the analysis because U_i is neither monotone nor convex in γ. Standard tools from convex optimization fail to apply. In addition, the solution to γ does not separate the two purposes of the escalation rule. It is hard to interpret and fails to isolate the economic trade-off of arbitration.

To isolate that trade-off and to facilitate the analysis we propose a dual problem to (P_{min}). Its objective consists of two elements. A measure on welfare and a measure on discrimination. Both measures are *on the continuation game.* We state the dual problem here and postpone further discussion until later.

The measure on welfare is the expected continuation value given a public information structure $\mathcal{B}$. Any public information structure implicitly defines a distribution of types in the continuation game. Taking expectations over types and aggregating over disputants we define the average expected utility in the continuation game by $\mathbb{E}[U|\mathcal{B}]$.

To construct the measure on discrimination define the difference in continuation utilities of a strong type and a weak type *pretending to be strong* as $D_i(\mathcal{B}) := U_i(1; 1, \mathcal{B}) - U_i(1; K, \mathcal{B})$. We refer to this term as the *ability premium* of the strong type. If we weight the ability premium by the inverse hazard rate we obtain the virtual rent $\Psi_i(1; \mathcal{B}) := D_i(\mathcal{B})(1-p)/p$. Weak types have no ability premium and thus $\Psi_i(K; \mathcal{B}) \equiv 0$. Taking expectations over types and aggregating results in $\mathbb{E}[\Psi|\mathcal{B}]$. The dual problem is

$$\begin{aligned} \max_{\mathcal{B}} \quad & \mathbb{E}[U|\mathcal{B}] + \mathbb{E}[\Psi|\mathcal{B}] \text{ s.t.} \\ & b_i(1) \geq p, \\ & \mathcal{B} \text{ consistent with } p. \end{aligned} \tag{P_{max}}$$

In Section 4 we show that this duality holds for a general class of escalation games. Next, we state the main properties of the solution to the dual problem. A full characterization of the optimal information structure and thus the optimal mechanism is part of the proof of Proposition 1 in appendix C.

Proposition 1. *Optimal alternative dispute resolution has the following features.*

(Asymmetry). *The information each disputant obtains within ADR differs with the disputant's identity, $b_A(\theta) \neq b_B(\theta)$.*

(No information trading). *The information each disputant obtains within ADR is independent of her own behavior, $b_i(1) = b_i(K)$.*

(Any match can escalate). *Any match escalates to litigation with positive probability, $b_i(\theta) \in (0, 1)$.*

Moreover, no type vetoes the mechanism with positive probability.

The main intuition behind Proposition 1 is that ADR should provide information so that all types find it beneficial to undergo the process of ADR even when being optimistic about litigation outcomes. At the same time, ADR should not provide too much information. Otherwise some types find it beneficial to abuse ADR to extract information only. As a result, disputants should receive the same information independent of their behavior during ADR. We address all properties in turns.

Asymmetry. A well-known feature of contests is that asymmetries decrease aggregate investment. In our setting that means, the more asymmetric litigation the smaller the expected quality of evidence provision. Increasing the quality of evidence provision is costly and thus any such reduction is beneficial to aggregate expected welfare in litigation. If litigation happens with positive probability, that reduction also increases expected welfare from participating in ADR. Larger welfare from participation eases participation constraints, and is thus beneficial to the arbitrator.

The stronger the asymmetry the larger the settlement share that ADR has to promise the party that is disadvantaged in the continuation game. The promise is costly to the arbitrator and she optimally implements some interior degree of asymmetry.

No Information Trading. No information trading implies that disputants cannot influence how much they learn about their opponent via their behavior during ADR. The no-information-trading condition is relevant because disputants gain no information advantage by mimicking under no information trading. The effect becomes immediate when considering Lemma 1 and 2 and the associated Figures 1, and 2.

If there is information trading, deviations alter the conditional distribution of types. The non-deviator is not surprised by escalation and acts *as if* on-path. She best responds to her *on-path mindset*, $\mathcal{B}$, and is not aware of the true distribution of types. The deviator, instead, is aware of her own deviation and holds correct beliefs. She is no longer indifferent and puts full mass on a single action.

Reducing the information advantage ceteris paribus reduces a deviator's utility, and decreases information rents to weak disputants. At $b_i(1) = b_i(K)$ the information advantage is minimized and equal to 0. Reducing the information advantage is of first-order importance to the arbitrator because actions in litigation are sensitive to beliefs.

The second result resembles the intuition from a second-price auction. There, to ensure incentive compatibility, the payment conditional on winning is independent of a bidder's type report. Similarly here, to ensure incentive compatibility, the belief conditional on escalation is independent of a disputant's type report.

Any match can escalate. The property implies that no "easy settlements" exist. Suppose–to the contrary–that the arbitrator guarantees settlement if both disputants are weak. Further, assume that both A and B are weak, but A mimics the strong type in ADR. If B observes escalation, she is sure to face a strong disputant A. She is pessimistic about her chances of winning litigation. The pessimism discourages her to invest into evidence provision. Disputant A can leverage on B's pessimism. A has to invest little into evidence to win against a weak B simply because B *expects* A to be strong.

The arbitrator has two instruments to make these deviations less attractive. First, she can induce escalation also for a pair of weak types. Weak types become less pessimistic and increase the quality of evidence. Second, the arbitrator can increase the likelihood that two strong types face each other in litigation. After mimicking the strong, a weak disputant faces a strong opponent in litigation more often. Moreover, that strong opponent expects a stronger opponent and increases investment. Both effects make deviations less profitable.

Both instruments ease incentive constraints and have an adverse effect on the strong types' participation constraints. However, increasing weak types' evidence quality is less harmful than increasing the share of strong types from a strong types' perspective. Thus, the first alternative has a smaller effect on participation constraints. As a result, any match escalates with positive probability.

Additional Information Revelation through the Arbitrator. So far we focused on the *implicit signal* that the decision to escalate implies. In addition, the arbitrator can release an *explicit signal* to influence disputants behavior further. Usually, arbitrators are not allowed to privately disclose the information obtained by one of the disputants to the other disputant. Public statements, however, are permitted. We thus focus on public signals in our discussion here.

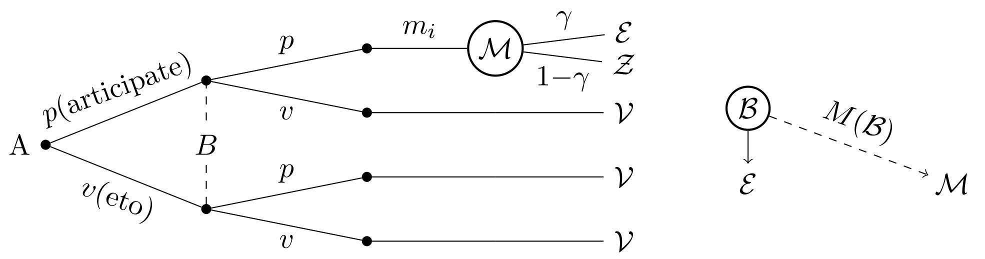
It is well-known from the literature on Bayesian persuasion (Kamenica and Gentzkow, 2011) that the scope for additional signals is to 'concavify' the problem (Aumann and Maschler, 1995). Our belief-management problem is a maximization problem and therefore its unconstrained optimum is on the concave closure of the objective with respect to the choice set. Thus, signals can only improve ADR if the ignored constraint $b_i(1) \geq p$ is violated at the unconstrained optimum. The constraint is violated if and only if $p > 1/3$. Indeed, the arbitrator provides a signal in such a case.

However, even if $p > 1/3$, the constraint binds at most for one disputant, A or B, but never for both. The following *symmetrizing signal* is optimal. Fix any mechanism $\mathcal{M}$ with $b_A \geq b_B$. After disputants report their types, the arbitrator performs a public and unbiased coin-flip. If the coin shows heads, she implements $\mathcal{M}$. If the coin shows tails, she relabels disputants (A becomes B and vice versa). Thereafter she implements $\mathcal{M}$. If $\mathcal{M}$ can be implemented, so can its symmetrized version because the coin-flip happens after reporting. Moreover, under the symmetrized version of unconstrained optimal ADR, the strong types' incentive constraints are satisfied. We state a stronger result showing that the arbitrator wants to send at most the symmetrizing signal.

Proposition 2. *The optimal public signal is the symmetrizing signal. The optimal direct mechanism must involve a public signal if and only if $p > 1/3$.*

2.4 The Economics of Belief Management

In this part we discuss the derivation and the economic interpretation of the dual problem belief management. First, we show how belief management highlights the economic trade-off the arbitrator faces. Second, we discuss the link between belief management and the



(a) Mechanism-Design Representation.

(b) Belief-Management Representation.

Figure 3: *Arbitration as Mechanism-Design Problem vs. Arbitration as Belief-Management Problem.* The mechanism-design problem considers behavior before, during, and after the mechanism. It affects the participation stage and thus all possible continuation events: veto, $\mathcal{V}$, settlement, $\mathcal{Z}$, and the on-path escalation game, $\mathcal{E}$. Belief management restricts attention to the analysis of the on-path escalation game, $\mathcal{E}$. A function M translates any implementable information structure $\mathcal{B}$ into a unique candidate mechanism.

initial mechanism-design problem. We show that similar to Grossman and Hart (1983) the belief-management dual disentangles cost of eliciting information from the benefits of (re-)distributing information. Using the function that links cost and benefits, we translate the properties of Proposition 1 back into properties of the arbitrator's choices in the mechanism-design problem.

The Belief-Management Approach. Belief management allows us to identify the economic trade-off. First, the optimal mechanism *reduces inefficiencies* in litigation if it cannot achieve settlement. All else equal, *everyone* benefits more from participating in ADR if subsequent litigation is less costly. Expected welfare is

$$\mathbb{E}[U|\mathcal{B}] := \sum_i \Big(\rho_i U_i(1; 1, \mathcal{B}) + (1-\rho_i) U_i(K; K, \mathcal{B})\Big),$$

where $\rho_i := b_{-i}(1)b_{-i}(K)/(b_{-i}(1)(1-b_i(1)) + b_i(1)b_{-i}(K))$ describes the likelihood that i is strong conditional on reaching the escalation game.

Second, the optimal mechanism induces an escalation game that is *fundamentally discriminatory.* That is, it aims at increasing the distance in expected continuation utilities between a true strong disputant and a weak disputant that *pretended to be strong* during ADR. Both face the same distribution of opponents, with the pretending disputant having an additional information advantage.

Aggregate expected discrimination given information structure $\mathcal{B}$ depends both on on-path best-response functions and on off-path best-response functions. Given the on-path best response a_i^*, a strong disputant's advantage over a weak disputant pretending to be strong is her ability premium. With abuse of notation let $U_i(m_i, a_i; \theta_i, \mathcal{B})$ be i's expected utility given action a_i. Let a_i^D be the deviator's best response. We decompose the ability premium into the difference of two terms,

$$D_i(\mathcal{B}) = \underbrace{U_i(1, a_i^*(1); 1, \mathcal{B}) - U_i(1, a_i^*(1); K, \mathcal{B})}_{\text{fundamental difference } (\geq 0)} - \underbrace{\Big(U_i(1, a_i^D(K); K, \mathcal{B}) - U_i(1, a_i^*(1); K, \mathcal{B})\Big)}_{\text{information advantage}(\geq 0)}.$$

The first term is a *fundamental difference* due to different cost functions. The second term is the deviator's *information advantage*. This term reduces the ability premium but is of smaller magnitude than the fundamental difference. The information advantage results from the deviator's superior information and her adjustment of the litigation strategy. To deter deviation, the arbitrator reduces the information advantage. The virtual rent, Ψ_i, weights the ability premium by the inverse hazard rate. The expected virtual rent, $\mathbb{E}[\Psi|\mathcal{B}]$, is a measure of discrimination.

The dual objective contains all constraints from problem (P_{min}) but the strong types' incentive constraints. The weak types' settlement value is pinned down through her binding incentive constraint. We substitute for settlement values in the strong types' incentive constraints, apply Bayes' rule and rearrange. We obtain the belief-management formulation of the strong types' incentive constraints,

$$\frac{b_{-i}(1)}{p} \underbrace{\left(U_i(1; 1, \mathcal{B}) - U_i(1; K, \mathcal{B})\right)}_{D_i(\mathcal{B})} - \frac{1 - b_{-i}(1)}{1 - p} \Big(U_i(1; K, \mathcal{B}) - U_i(K; K, \mathcal{B})\Big) \geq 0 \quad \forall i. \quad (3)$$

In our stylized model (3) reduces to $b_i \geq p$. Both the objective and its constraints are formulated on the escalation game directly. That reduces the problem's complexity.

Belief Management and the Optimal Mechanism. The optimal information structure is unique up to a symmetric switch between disputants. The optimal mechanism, however, is not unique. The reason is that there is some degree of freedom in the sharing rule x_i. Recall that we can represent the expected utility from participating in ADR by the settlement value and the escalation value alone. That is, $\Pi_i(m_i, \theta_i) = z_i(m_i) + y_i(m_i; \theta_i)$. The escalation value $y_i(m_i, \theta_i)$ is entirely determined by γ_i, while the settlement value $z_i(m_i)$ is jointly determined by x_i and γ_i. We take a reduced-form approach (Border, 2007) and identify a reduced-form mechanism by (z, γ). The reduced-form mechanism is unique.

On the reduced-form level, the two problems (P_{min}) and (P_{max}) are related through a mapping, $(z, \gamma) = M(\mathcal{B})$. The mapping links every information structure to a unique *candidate mechanism*. The arbitrator bears a cost of implementing a given $\mathcal{B}$. It is the likelihood of escalation necessary to implement $\mathcal{B}$ in an incentive-compatible way. The mapping M determines the lowest such cost for any given $\mathcal{B}$. The belief-management objective, in turn, determines how $\mathcal{B}$ performs in terms of efficiency and discrimination. The objective follows by substituting the binding constraints into the objective of (P_{min}).

Given the results from Proposition 1, we use $M(\mathcal{B})$ to determine the optimal mechanism. Figure 4 depicts the values of z_i and y_i as well as the escalation rule, γ_i. More

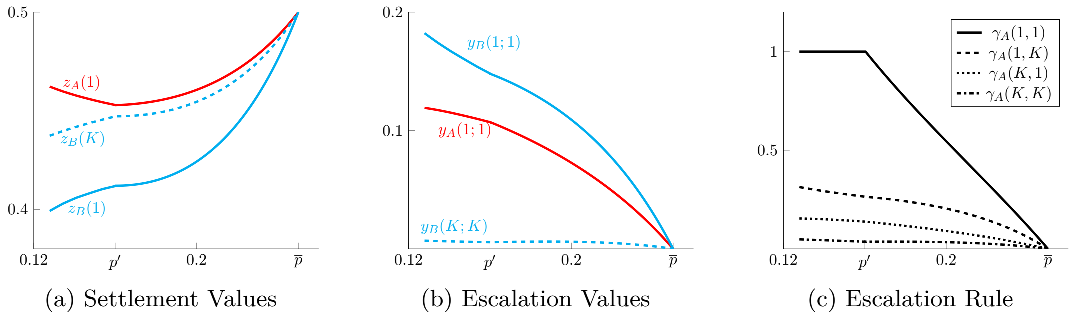


Figure 4: Properties of the optimal mechanism as a function of p (with $K = 3$). To the left of p' no interior solution exists and instead a boundary solution with $\gamma(1,1) = 1$ is optimal. To the right of $\overline{p}$ full settlement is possible. In any case, $z_A(K) = z_A(1)$ and $U_A(K; K, \mathcal{B}) = 0 \Rightarrow y_A(K,K) = 0$.

generally, we obtain the following properties.

Corollary 1. *Litigation after a failed attempt of optimal ADR is less inefficient than litigation under the prior. Furthermore, optimal ADR has the following properties.*

(i) Strong types escalate more often, $\gamma_i(1) \geq \gamma_i(K)$*, and* $\gamma_B(1) > \gamma_B(K)$*.*

(ii) Disputant A expects a higher settlement value, $z_A(1) = z_A(K) > z_B(K) > z_B(1)$*.*

(iii) Disputant B expects a higher utility in litigation, $U_i(1; 1, \mathcal{B}) > U_B(K; K, \mathcal{B}) > U_A(K; K, \mathcal{B}) = 0$*.*

(iv) A is weakly better-off than B, i.e., $\Pi_A \geq \Pi_B$*. If the optimal belief system is in the interior of the feasible set, both expect the same payoff from participation.*

In appendix C we present and discuss further properties of the optimal mechanism, including analytic solutions for parameter ranges where they exist.[8]

2.5 Relationship to the Conflict-Resolution Literature

In this part we relate our approach to both the existing literature on second-best arbitration mechanisms and the classical bilateral bargaining approach.

Other (Second-Best) Mechanisms. Formally, the existing literature focuses on constant-sum escalation games that have an ex-post equilibrium. These models build on Bester and Wärneryd (2006) and are widely applied to international conflicts. Recently a series of papers studies how arbitration mechanisms interact with different degrees of uncertainty (Fey and Ramsay, 2011), ex-post veto rights (Hörner, Morelli, and Squintani, 2015), and strategic militarization prior to the conflict (Meirowitz et al., 2017).

Net of the escalation game, the model of arbitration in these papers is very close to ours. Thus, we can apply the belief-management approach to these models immediately. Results differ fundamentally.

First, in constant-sum games, aggregate welfare conditional on escalation, $\mathbb{E}[U|\mathcal{B}]$, is invariant to the information structure. Second, ex-post equilibrium implies that informa-

[8] A matlab program calculating the optimal solution for any set of parameters (p, K) is available on the authors' websites.

tion revelation has no effect on behavior. Type θ_i's optimal action, $a_i^{\theta_i}$, does not depend on the information structure. Thus, $U_i(m_i; \theta_i; \mathcal{B})$ varies only—and linearly—through $b_i(m_i)$. There is no loss in considering (ex-post) symmetric mechanisms only, and only the fundamental difference matters. The difference to the results from Proposition 1 is immediate when considering its equivalent in Hörner, Morelli, and Squintani (2015).

Proposition 3. *Optimal arbitration in the model of Hörner, Morelli, and Squintani (2015) has the following features.*

(Symmetry) *The information disputants obtain is independent of their identity, $b_A{=}b_B$.*

(Information trading). *The information a disputant obtains depends on her behavior in the mechanism, $b_i(1) \neq b_i(K)$.*

(Weak types settle). *Two weak types always settle, and thus $b_i(K) = 1$.*

Proposition 3 is the belief-management version of the arbitration result in Hörner, Morelli, and Squintani (2015). The results in Proposition 3 oppose those from Proposition 1. This highlights the important role the escalation game has on optimal arbitration. We prove a general version of Proposition 3 for escalation games with ex-post equilibrium in the supplementary appendix E.

Hörner, Morelli, and Squintani (2015) obtain a *sorting mechanism.* "Weak enough" matches enjoy guaranteed settlement. Intermediate matches sometimes escalate and sometimes settle. "Strong enough" matches are guaranteed to escalate.

Proposition 1 demonstrates that an effect of information on behavior in the escalation game overturns that results. Once ADR fails, disputants reason about the cause of escalation and adapt their continuation strategies accordingly. Noiseless contests are particularly sensitive to changes in the information structure. The change in behavior becomes the primary concern of the arbitrator. It leads to the results from Proposition 1.

When can we expect information-sensitive escalation games? The effect on continuation strategies is important if disputants have sufficient time to react to the information they obtain within arbitration. Adjusting strategies may be difficult in international conflicts. Failure of settlement negotiation may immediately lead to war, leaving disputants no time to re-optimize military strategies. Strategies are only functions of the information prior to arbitration and the continuation game has an ex-post equilibrium.

Legal disputes are different. Disputants face a sufficient time lag between failed settlement and the beginning of formal litigation.[9] That time lag allows for adjustments on litigation strategies.

Pretrial Negotiations and Simple Bargains. A seminal strand of literature following Bebchuk (1984) considers bilateral and unmediated bargaining processes in the shadow of litigation.[10] Brown and Ayres (1994) point out that managing information flow between

[9] Litigation follows a strict procedure overseen by the court. Courts typically do not have excess capacities which leads to long waiting times between failed ADR and litigation. Consequently, parties can adjust strategies before entering formal litigation.

[10] There is a large literature in that tradition, see Spier (2007) for an overview and Vasserman and Yildiz (2018) for a recent contribution.

disputants through a third party can be beneficial. In line with that finding our mechanism performs strictly better than bilateral bargaining.

Similar arguments apply to court-proposed settlements in countries with an inquisitorial law system. The court proposes a particular result, a *simple bargain* (x_A, x_B), which may be accepted in some cases and rejected in others. A feature of these simple bargains is that only a veto of either disputant triggers escalation. As a result, types are only relevant under escalation, and it is without loss of generality to make (x_A, x_B) independent of the disputant's type.[11]

Simple bargains significantly restrict the set of feasible information structures compared to general mechanisms. Consider, for example, the situation in which only A rejects. Since B knows the veto was A's decision, B's post-veto belief about A cannot depend on her own type nor on her decision. In turn, A, who is aware that she triggered the veto, can base her updating only on B's decision not to veto and not on her own action choice. Moreover, veto decisions have to be *optimal*. Disputant i only vetoes if a veto yields higher continuation utility than ratification.

In the ADR model above simple bargains are optimal if and only if full settlement can be guaranteed. In that case, a veto never occurs and the mechanism becomes a simple bargain. In all other cases, our full-support result contradicts that implementation.[12] Weak types strictly benefit from ADR. They never veto ADR. Full-support requires occasional escalation also for the match (K, K). A contradiction.

The comparison between ADR and simple bargains emphasizes the relation between information design and mechanism design. The arbitrator's commitment to conceal some information once omniscient, reduces her cost of eliciting information. Disputants are willing to provide information if not all of it becomes public. In simple bargains, in turn, information revelation happens via *observed action choices* and is thus public by design. Naturally, this restriction increases the price of revelation and harms settlement rates.

3 General Model

In this section we describe our general model. We allow for an arbitrary number of types, a large class of escalation games, and differences between escalation after a veto and escalation after unsuccessful arbitration.

Grand Game. Two ex-ante identical, risk-neutral players have a conflict over a pie worth 1 to each. By default, they solve the conflict via a given game, $\mathcal{V}$. Before playing $\mathcal{V}$, players can agree to an arbitration mechanism, $\mathcal{M}$.

Arbitration. Arbitration is a mechanism proposed by a non-strategic third party, the arbitrator, at the beginning of the grand game. It leads either to settlement, $\mathcal{Z}$, or to

[11] Strulovici and Siegel (2018) provide an economic foundation for a similar mechanisms in adversarial law systems in the context of plea bargaining.

[12] In a recent paper Zheng and Kamranzadeh (2018) look at simple bargains in a setup identical to the one in this section. Indeed, the "optimal split" they derive is strictly inferior to the optimal mechanism.

escalation, $\mathcal{E}$. Settlement awards player i a share of the pie, x_i, such that $x_A + x_B \leq 1$. Escalation triggers a non-cooperative game $\mathcal{E}$. The two games $\mathcal{E}$ and $\mathcal{V}$ may be identical but do not have to be.[13] In any case, $\mathcal{E}$, too, is beyond the arbitrator's control. That is, once the conflict escalates the arbitrator controls neither players' action choices nor the rules of $\mathcal{E}$. Arbitration thus describes a mechanism-design problem within a greater strategic environment beyond the arbitrator's control. The structure sketched in Figure 3 on page 14 applies to the general model, too.

Initial Information Structure. Each player i has a private type, θ_i, independently drawn from the same distribution over $\Theta = \{1, 2, ..., K\}$. The probability of being type θ_i is $p(\theta_i) > 0$. Types are payoff relevant in $\mathcal{V}$ and $\mathcal{E}$ but not under settlement. Types order players' cost functions from low ($\theta_i = 1$) to high ($\theta_i = K$) in both $\mathcal{V}$ and $\mathcal{E}$. We assume that payoffs in both games are monotone decreasing in types (see below): the lower (higher) the type, the stronger (weaker) the player. All but the realization of types is common knowledge.

Payoffs. Under settlement, each player receives a payoff equal to her share of the pie, x_i. Conditional on escalation, payoffs depend on the play of the escalation game $\mathcal{E}$. Let $A^{\mathcal{E}} \subset \mathbb{R}^2$ describe the space of joint actions in $\mathcal{E}$. We assume that fights are never efficient. That is, the payoff function of the escalation game $\mathcal{E}$ is a mapping $(u_A, u_B) : \Theta^2 \times A^{\mathcal{E}} \to (-\infty, 1]^2$ with $u_A + u_B \leq 1$. Similarly, the payoff function of the veto game $\mathcal{V}$ is $(v_1, v_2) : \Theta^2 \times A^{\mathcal{V}} \to (-\infty, 1]^2$ with $v_A + v_B \leq 1$.

Both u_i and v_i are weakly decreasing in θ_i, and symmetric. That is, $u_i \equiv u$ and $v_i \equiv v$.

Timing. The timing is identical to the stylized model. First, players privately observe their types and the arbitrator announces $\mathcal{M}$. Then, they simultaneously decide whether to participate in or to veto arbitration. Upon veto, players learn who vetoed $\mathcal{M}$, play $\mathcal{V}$, and payoffs realize. If both participate in $\mathcal{M}$, they play $\mathcal{M}$ which either leads to settlement or to escalation. Settlement implements x_i, escalation triggers updating of beliefs and the play of the escalation game $\mathcal{E}$ under these beliefs.

Solution Concept. We use perfect Bayesian equilibrium (Fudenberg and Tirole, 1988). We look for a mechanism and an associated equilibrium to maximize the probability of settlement. We discuss alternative objectives in Section 5.

Key Modeling Choices. We formally analyze arbitration, that is, a mechanisms with interim participation constraints. Our results, however, extend to conflict-management mechanisms in general.

A key feature of our model is that players differ in their ability to play the escalation game. For simplicity, we assume that types are irrelevant under settlement. Escalation is the arbitrator's only screening device. Results extend to richer settings, if types can be ordered. We discuss them in Section 5 and the supplementary appendix J.

A second important assumption is that settlements are based on soft information only. In reality, such mechanisms are common as they are the least-costly solutions. Nev-

[13] In the legal context the rules of formal litigation after a failed attempt of ADR may differ from those that prevail if ADR is rejected altogether. See Prescott, Spier, and Yoon (2014) for a specific example.

ertheless, our setting nests more complicated mechanisms that involve on-path screening. Incorporating further screening options is straightforward. They enter as properties of the escalation game $\mathcal{E}$. Our mechanism describes the initial exchange of soft information.

Finally, the assumption that the arbitrator can ex-ante commit to her protocol is in line with the vast majority of the mechanism-design literature, but nevertheless restrictive. We ignore important cases in which renegotiation *through the arbitrator* is anticipated by parties in advance. Real-world arbitrators mostly provide their services repeatedly which incentivizes them to commit.

4 Analysis

4.1 First-Best Benchmark: Full-Settlement Solutions

We first provide conditions for a full-settlement mechanism. Types are payoff irrelevant under settlement. Thus, any mechanism avoiding both $\mathcal{V}$ and $\mathcal{E}$ is a pooling mechanism.

Any pooling mechanism implements a particular sharing rule (x_A, x_B) independent of the players' reports. Whether such a rule satisfies the participation constraints depends on strategic choices after a veto. A player's optimal strategy in $\mathcal{V}$ maximizes her payoff over action choices conditional on her updated belief.

Suppose player i vetoes the mechanism. She learns nothing about her opponent from her own veto decision and believes that $-i$'s type is distributed according to the prior p. The non-vetoing player $-i$ may change her beliefs after i's veto. Her updated distribution over i's type is ρ^V which in a full-participating equilibrium is an off-path belief. By the "no-signaling-what-you-don't-know" condition ρ^V is independent of θ_{-i}.The public information structure after i vetoes is (p, ρ^V). Behavior in the continuation game follows a Bayes Nash equilibrium given (p, ρ^V). Player i's expected continuation payoff is

$$V_i(\theta_i, (p, \rho^V)) := \max_{a_i \in A_i^{\mathcal{V}}} \sum_{\theta_{-i}} p(\theta_{-i}) \int_{A_{-i}^{\mathcal{V}}} v(\theta_i, \theta_{-i}, a_i, a_{-i}) \mathrm{d}F(a_{-i}|\theta_{-i}, (p, \rho^V)), \qquad \text{(V)}$$

where $F(a_{-i}|\theta_{-i}, (p, \rho^V))$ is the conditional distribution over equilibrium action choices of θ_{-i} given information structure (p, ρ^V). Our first result is in the spirit of Compte and Jehiel (2009) and Zheng (2018). It determines a necessary and sufficient condition for a full-settlement mechanism.

Proposition 4 (Full Settlement Mechanisms)**.** *Optimal arbitration guarantees full settlement if and only if there is a distribution ρ^V such that $V_i(1, (p, \rho^V)) \leq 1/2$.*

The result combines three constraints. (i) Settlement has no screening power, thus the arbitrator offers a pooling mechanism; (ii) the arbitrator is budget constrained, thus settlement divides the pie; and (iii) participation is voluntary, thus the arbitrator incentivizes players via a sufficiently large share. Players are ex-ante symmetric, and a first-best mechanism exists only if $\mathcal{V}$ is sufficiently costly. Proposition 4 reduces existence of the first-best mechanism to the simple question: *Can we implement an equal split of the pie?*

4.2 Second-Best Mechanisms: The Arbitrator's Problem

We now analyze the second-best mechanism. We impose a set of assumptions on the veto game $\mathcal{V}$ to facilitate the analysis. The set of functions V_i describes a reduced form of $\mathcal{V}$ and enters the arbitrator's problem as a primitive. The arbitrator has no direct control over the players' decisions, given (p, ρ^V). As we know from Section 2, understanding the continuation games is necessary to solve the design questions. Hence we make our (technical) assumptions directly on the functional form of V_i. Let $conv_x f(t,x)$ be the largest function weakly smaller than $f(t,x)$ and convex in x.

Assumption 1. V_i exists for all (p, ρ^V), is non-negative, and satisfies the following.

(HC) Upper hemicontinuity. V_i is upper hemicontinuous in (p, ρ^V).

(S) Symmetry. $V_A(\theta, (p, \rho^V)) = V_B(\theta, (\rho^V, p))$ for any θ, (p, ρ^V).

(OST) Optimistic strongest type. $V_i(1, (p, \rho^V)) > 1/2$ for any ρ^V.

(CONV) Convex envelope. $V_i(\theta, (p, \rho^V)) = conv_p V_i(\theta, (p, \rho^V))$ for any θ, ρ^V.

Non-negativity implies that there always is at least an option to concede, that is, giving up on winning the pie at no additional cost. Property (HC) implies that the arbitrator's objective is continuous in her choices. (S) assumes a symmetric, anonymous equilibrium. (HC) guarantees existence of an optimum and (S) significantly reduces the notational burden. (OST) ensures that Proposition 4 does not apply. We assume the technical property (CONV) for convenience. Together with Assumption 2 made below it allows us to focus on full-participation mechanisms. We discuss alternatives in Section 5. Focussing on settings with full-participation simplifies the analysis in two ways. First, it reduces the cases to be considered. Second, veto-beliefs ρ_i^V are off-path and therefore arbitrary. We can restrict attention to the on-path belief structure.

Second-best arbitration uses escalation as a screening device. Once the conflict escalates, the arbitrator's influence ceases and $\mathcal{E}$ is played non-cooperatively. We focus on direct revelation mechanisms. The vector of type reports maps into (i) a probability that the conflict escalates, γ, (ii) a sharing rule under settlement, X, and (iii) an additional public signal, Σ, with realization σ. That is, the mechanism is a mapping

$$\mathcal{M}(\cdot) = (\gamma(\cdot), X(\cdot), \Sigma(\cdot)) : \Theta^2 \to [0,1] \times [0,1]^2 \times \Delta(\mathscr{S}), \tag{$\mathcal{M}$}$$

where $\Delta(\mathscr{S})$ is the set of probability distributions over an arbitrary countable set $\mathscr{S}$. Although public signals are necessary to invoke the revelation principle, the effect of the implicit signal sent via γ is sufficient to describe the economic intuition. In the main text, we suppress signals notationally whenever convenient to keep the exposition simple. All our results include the choice of the signaling function. We focus on public signals in the main part and extend our results to private signals in Section 5.[14]

Let $\Pi_i(m_i; \theta_i)$ be the (continuation) value from participating and reporting m_i in a

[14] Although we impose the restriction to public signals, our setting nests all game forms with observable actions including all forms of decentralized and bilateral negotiation.

given mechanism. The ex-ante probability of escalation is $Pr(\mathcal{E})$. Conditional on full participation the arbitrator solves the same program as in Section 2.

$$\begin{aligned} &\min_{(\gamma,X,\Sigma)} Pr(\mathcal{E}) \text{ s.t.} \quad \forall i, m_i, \theta_i : \\ &\Pi_i(\theta_i;\theta_i) \geq \Pi_i(m_i;\theta_i), \\ &\Pi_i(\theta_i;\theta_i) \geq V_i(\theta_i,(p,\rho^V)). \end{aligned} \tag{P$_{min}$}$$

The first set of constraints ensures incentive compatibility, the second full participation. The veto belief ρ^V is arbitrary since vetoes happen off the equilibrium path. The value from participation, Π_i assumes optimal behavior in the continuation game.

4.3 Incentives and Information

There is a conceptual difference between the information sets in games $\mathcal{V}$ and $\mathcal{E}$. Different to the veto game, players have a private history of play, m_i, at the beginning of $\mathcal{E}$. The public information structure, $\mathcal{B}$, specifies a belief for any on-path private history, that is, for any possible type report θ_i. It is a $K \times 2$ matrix of probability distributions. An element of this matrix, $\beta_i(\cdot|\theta_i)$, is θ_i's *individual belief*, her on-path expectations about $-i$'s type distribution. The K-dimensional vector, $\boldsymbol{\beta}_i$, collects i's possible beliefs, and $\mathcal{B}:=(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B)$. The definition of the continuation utility extends naturally.

$$U_i(m_i;\theta_i,\mathcal{B}) := \sup_{a_i \in A_i^{\mathcal{E}}} \sum_{\theta_{-i}} \beta_i(\theta_{-i}|m_i) \int_{A_{-i}^{\mathcal{E}}} u(\theta_i,\theta_{-i},a_i,a_{-i}) \mathrm{d}G(a_{-i}|\theta_{-i},\mathcal{B}), \tag{U}$$

where $G(a_{-i}|\theta_{-i},\mathcal{B})$ is θ_{-i}'s distribution over equilibrium actions.[15] We impose simplifying assumptions on U_i for notational convenience.

Assumption 2. *U_i exists, is non-negative, and satisfies (HC) and (S) for all $\mathcal{B}$. In addition, for any information sets such that both U_i and V_i are defined, $U_i \geq V_i$.*[16]

The last part of Assumption 2 implies that players are not *exogenously* punished for participating in arbitration. Together with Assumption 1 it is sufficient for full participation at the optimum. Consequently, vetoes are off-path events.

The continuation payoff (U) illustrates the main features of our model. First, individual beliefs *depend on type reports only.* Second, action choices depend on the entire belief system, $\mathcal{B}$. On the equilibrium path players expect their opponents to report truthfully. In addition, they expect them to follow their equilibrium strategies given $\mathcal{B}$ in the escalation game. All elements in (U) but $\mathcal{B}$ are beyond the arbitrator's control, in particular the functional form of U_i. Therefore, we treat U_i as a primitive to the arbitrator's problem.

Assume no additional public signal is sent. The (continuation) value from participation

[15] We assume that an equilibrium exists in any game so that whenever $m_i = \theta_i$ a maximum exists. However, a maximum may not exist if $m_i \neq \theta_i$ which is why we choose the more flexible supremum.

[16] Formally, define the projection $I^{\mathcal{E}}(I^{\mathcal{V}}) := \{(m_i;\theta_i,\mathcal{B})|m_i{=}\theta_i, \mathcal{B}{=}(\rho_i,\rho_{-i})_K, \text{ and } (\theta_i,(\rho_i,\rho_{-i})) = I^{\mathcal{V}}\}$, with $(\rho_i,\rho_{-i})_K$ a $K \times 2$-matrix such that each row equals to (ρ_i,ρ_{-i}). Then, $U_i(I^{\mathcal{E}}(I^{\mathcal{V}})) \geq V_i(I^{\mathcal{V}})$.

in the general model becomes

$$\Pi_i(m_i;\theta_i) = \underbrace{\sum_{\theta_{-i}} p(\theta_{-i})(1-\gamma_i(m_i,\theta_{-i}))x_i(m_i,\theta_{-i})}_{=:z_i(m_i)(\text{settlement value})} + \underbrace{\gamma_i(m_i)U_i(m_i;\theta_i,\mathcal{B})}_{=:y_i(m_i;\theta_i)(\text{escalation value})}, \tag{4}$$

with $\gamma_i(m_i) := \sum_{\theta_{-i}} p(\theta_{-i})\gamma_i(m_i,\theta_{-i})$ the expected escalation probability. It follows that

$$\Pi_i(\theta_i;\theta_i) - \Pi_i(\theta_i;\theta_i{+}1) = \gamma_i(\theta_i)(\underbrace{U_i(\theta_i;\theta_i,\mathcal{B}) - U_i(\theta_i;\theta_i{+}1,\mathcal{B})}_{=:D_i(\theta_i;\theta_i,\mathcal{B})}).$$

Any choice of γ has a non-linear effect on U_i via $\mathcal{B}$. Standard methods are not directly applicable. We proceed with a change of variable.

4.4 Belief Management

In this part we show that the choice of the optimal post-escalation belief system is sufficient to determine the optimal mechanism. Thus, we reduce the problems' dimensionality and transform it into a problem of managing beliefs in the escalation game. The transformation not only reduces complexity, but it also separates the mechanism-design part of eliciting information from the information-design part of distributing that information. We later use the belief-management approach to characterize the optimal mechanism via the properties of the escalation game $\mathcal{E}$.

Two steps yield our result. First, we determine the reduced form of a mechanism (Border, 2007). This step allows us to work with expected shares, rather than ex-post shares. Second, we construct a one-to-one mapping between an information structure and a candidate for the optimal reduced-form mechanism.

Definition 1 (Reduced-Form Mechanism)**.** A tuple (z,γ) is the *reduced-form mechanism* of a mechanism (γ, X) if each element of $z_i \in z$ takes the functional form
$z_i(m_i) = \sum_{\theta_{-i}} p(\theta_{-i})(1-\gamma_i(m_i,\theta_{-i}))x_i(m_i,\theta_{-i})$.

We introduce two more concepts before stating our result. First, we define the set of *consistent belief systems* which describes all belief systems the arbitrator can implement given the prior. Consistency allows for $K+1$ independent belief distributions.[17]

Definition 2 (Consistency)**.** The set of *consistent belief systems*, $\{\mathcal{B}\}_p$, contains all $\mathcal{B}$ for which an escalation rule γ exist such that $\mathcal{B}$ follows from the prior and Bayes' rule.

Second, we extend consistency to settings with an additional public signal, Σ.

Definition 3 (Random Consistent Belief System)**.** A random variable, $\mathcal{B}(\Sigma)$, is a random consistent belief system, if it maps signal realization σ into consistent belief systems.

The random variable $\mathcal{B}(\Sigma)$ links a realization σ directly to a consistent belief system. If σ is induced with $Pr(\sigma)$ via Σ, so is the corresponding consistent belief system.

[17] Consistency corresponds to Bayes' plausibility in a Bayesian persuasion setting. We provide an in-depth discussion and a constructive characterization in the supplementary appendix I.

Any (stochastic) mechanism $\mathcal{M}$ trivially induces a random consistent belief system. Our first theorem shows that a similar statement holds in reverse. Any random consistent belief system, $\mathcal{B}(\Sigma)$, determines at most one candidate for the optimal (reduced-form) mechanism.

Theorem 1. *Under Assumption 1 and 2 the set of feasible random belief systems $\mathcal{B}(\Sigma)$ is compact. Moreover, for any feasible $\mathcal{B}(\Sigma)$ the optimal reduced-form mechanism, (z, γ), is unique.*

The proof of Theorem 1 is constructive. It characterizes the function $M : \mathcal{B}(\Sigma) \mapsto (z, \gamma)$ that identifies a *unique candidate* for any consistent belief system. The intuition behind the construction does not rely on the choice of the public signal, Σ. We suppress it notationally in the main text. We organize our discussion using a set of observations that correspond to the steps in the formal proof.

Observation 1. Belief system, $\mathcal{B}$, and continuation payoff, U_i, are homogeneous of degree 0 in the escalation rule. Escalation value, y_i, and escalation probability, γ_i, are homogeneous of degree 1 in the escalation rule.

Suppose player i submits a report m_i and learns that the conflict escalates. Then, the probability that she faces type $\tilde{\theta}_{-i}$ is $p(\tilde{\theta}_{-i})\gamma(m_i, \tilde{\theta}_{-i}) / \sum_{\theta_{-i}} p(\theta_{-i})\gamma(m_i, \theta_{-i})$ which is determined by the *relative* likelihood of escalation among the different possible *report profiles*. Thus, if γ implements $\mathcal{B}$ so does $\alpha\gamma$. The externality that γ imposes on the continuation payoff in $\mathcal{E}$ is entirely expressed by the belief system that γ induces. Thus, U_i is invariant to any scaling of the escalation rule. Finally, the probability of reaching escalation and hence the escalation value depend linearly on the escalation rule.

Observation 2. The *worst escalation rule* implementing a given belief system is unique.

Take any escalation rule that implements $\mathcal{B}$ 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 $\overline{g}_{\mathcal{B}} := \overline{\alpha}\gamma$ maximizes escalation and is the worst rule that implements $\mathcal{B}$. Identifying the worst escalation rule is sufficient to characterize *all escalation rules* implementing $\mathcal{B}$. The set of all γ implementing $\mathcal{B}$ is $\{\alpha\overline{g}_{\mathcal{B}} : \alpha \in (0, 1]\}$. Given $\mathcal{B}$, the problem reduces to finding the *lowest* α sucht that $\mathcal{M}$ satisfies the constraints in problem (P_{min}).

Observation 3. It is without loss to assume that any type θ_i with positive settlement value has either a binding incentive constraint, or a binding participation constraint in the second-best mechanism. Given $\mathcal{B}$, all constraints are linear in α and z.

The first part of the observation follows since no first-best mechanism exists. Full settlement fails because the arbitrator is budget constrained under settlement. Optimal arbitration ensures that some constraint binds for any type. The second part is immediate when combining a player's value from participating in the mechanism, expression (4), with Observation 2. Observation 3 implies that $\mathcal{B}$ captures the entire non-linear part of the constraints. Since each type has some binding constraint, the set of constraints consists of $2K$ independent linear equations with $2K+1$ unknowns, the $2K$ settlement values and the scalar α. To close the problem we need one more equation. We use the resource constraint of the arbitrator at the level of expected settlement payments,

$$\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \leq 1 - Pr(\mathcal{E}). \tag{5}$$

Observation 4. A reduced-form mechanism is feasible only if it satisfies (5).

The resource constraint, (5), is an immediate consequence of the arbitrator's budget constraint. The arbitrator can only allocate the pie if there is settlement. Thus, the total share allocated cannot be greater than the probability of settlement.

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. By Observation 1 the same holds for $1 - Pr(\mathcal{E})$. Since Proposition 4 does not apply, the resource constraint binds at the optimum and is the last equation needed. For any consistent $\mathcal{B}$ there is a unique tuple (z^*, α^*) satisfying the binding constraints with equality.

Via Observation 1 and 2, a feasible escalation rule implementing $\mathcal{B}$ exists if and only if the corresponding $\alpha^* \leq 1$. We partition the set of consistent belief systems into two subsets, $\{\mathcal{B}\}_\emptyset := \{\mathcal{B} |$ no feasible (z, γ) exists$\}$ and the remainder $\{\mathcal{B}\}_p \setminus \{\mathcal{B}\}_\emptyset$.

Finally, we construct a function $M : \mathcal{B} \mapsto (z, \gamma)$ that identifies a *unique candidate* (z, γ) for any $\mathcal{B} \notin \{\mathcal{B}\}_\emptyset$ and points to the origin otherwise. The function M is continuous in the interior of the support and given by

$$M(\mathcal{B}) := \begin{cases} (z^*, \alpha^* \overline{g}_{\mathcal{B}}) & , \text{ if } \mathcal{B} \notin \{\mathcal{B}\}_\emptyset \\ 0 & , \text{ if } \mathcal{B} \in \{\mathcal{B}\}_\emptyset. \end{cases}$$

Discussion of Theorem 1. The problem (P_{min}) contains both a mechanism-design part and an information-design part. The arbitrator decides on the game form and acts as a mechanism designer. Yet, under escalation her actions are restricted to distributing information. She acts as an information designer. However, any information she distributes she has to elicit first. The (promised) information distribution, in turn, influences the cost of information elicitation.

As main implication, Theorem 1 separates elicitation and distribution of information. For any belief system, the function M determines if that belief system satisfies the budget constraint. Moreover, M determines the escalation rule that minimizes the price (in terms of lost settlement shares) to elicit the necessary information.

Once the price of information is determined, the mechanism-design problem reduces to an information-design problem asking: *What is the optimal post-escalation information structure?*

4.5 A Dual Problem

What remains is to determine (i) when a reduced-form mechanism is implementable and (ii) how to find the optimal belief system. To address the first issue, we borrow the general implementation condition from Border (2007). The second issue depends, in general, on the details of the escalation game. Our characterization is constructive and focuses on what we call a *regular* environment. We address the two issues separately.

First, we provide a version of the results in Border (2007) to state necessary and sufficient conditions to implement a reduced-form mechanism (z, γ) via some $\mathcal{M}$. For any $Q \subset \Theta^2$ define $Q_i{:=}\{\theta_i | \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in Q\}$ and $\tilde{Q}{:=}\{(\theta_A, \theta_B) \in \Theta^2 | \theta_i \notin Q_i \text{ for } i{=}\{A, B\}\}$.

Proposition 5 (Border (2007))**.** *Take any reduced-form mechanism (z, γ). An ex-post feasible X that implements z exists if and only if*

$$\sum_i \sum_{\theta_i \in Q_i} z_i(\theta_i) p(\theta_i) \leq 1 - Pr(\mathcal{E}) - \sum_{(\theta_A, \theta_B) \in \tilde{Q}} (1 - \gamma(\theta_A, \theta_B)) p(\theta_A) p(\theta_B), \quad \forall Q \subseteq \Theta^2. \quad \text{(GI)}$$

Proposition 5 imposes a set of additional constraints. Following Border (2007) we call this set the general implementation condition, (GI).

Next, we restrict the environment to be *regular.* Regularity adds structure to the model as it determines precisely which constraints are *binding for sure* and which constraints are possibly *redundant.*

Regularity is based on two conditions. The first condition guarantees that the strongest type ex-ante expects to face a sufficiently weak opponent; that is, she is optimistic about her prospects. The second condition guarantees that a player's type, θ_i, becomes increasingly relevant (relative to the information structure) the weaker that player is. We generalize the ability premium from Section 2 to $D_i(m; \theta_i, \mathcal{B}) = U_i(m; \theta_i, \mathcal{B}) - U_i(m; \theta_i{+}1, \mathcal{B})$. We state that second condition in terms of difference ratios.

Definition 4 (MDR)**.** The game $\mathcal{E}$ 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.

Definition 5 (Regularity)**.** Let $\underline{\rho}^V := \arg\min_{\rho^V} V_i(1, (p, \rho^V))$. In a regular environment

(i) $\sum_i \sum_{\theta_i \in \hat{Q}} p(\theta_i) V_i(\theta_i, (p, \underline{\rho}^V)) < \sum_{\theta_i \in \hat{Q}} p(\theta_i)$, for any $\hat{Q} \subseteq \Theta$ and $\hat{Q} \neq \{1\}$, and

(ii) $\mathcal{E}$ satisfies (MDR).

Assumption 3 (Regularity)**.** The environment is regular.

Regularity, in particular (MDR), imposes strong restrictions on the nature of the escalation game. Yet, it allows us to focus on the economics when setting up the dual problem. We discuss below Theorem 2 and in Section 5 how the main insights carry over to a more general environment.

Lemma 3. *Under assumption Assumption 1 to 3 the following holds*

(i) The strongest type's participation constraint holds with equality. All other participation constraints are redundant.

(ii) Local downward incentive constraints hold with equality. The remaining downward incentive constraints are redundant.

Our final step towards the dual formulation is to extend the formulations of expected type distributions, (expected) virtual rents, expected aggregate welfare, and the set of constraints from those used in Section 2.

Definition 6 (Virtual Rent)**.** Player θ_i's virtual rent is $\Psi_i(\theta_i, \mathcal{B}) := w(\theta_i) D_i(\theta_i; \theta_i, \mathcal{B})$, with $w(\theta_i) := \Big(1 - \sum_{k=1}^{\theta_i} p(k)\Big) / \Big(p(\theta_i)\Big)$.

We are interested in the ex-ante type distributions conditional on escalation $(\rho_A(\cdot), \rho_B(\cdot))$. They are the solution to the linear system of equations

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

Furthermore, $\mathbb{E}[\Psi|\mathcal{B}] = \sum_i \sum_{\theta_i} \rho_i(\theta_i) \Psi_i(\theta_i, \mathcal{B})$, and $\mathbb{E}[U|\mathcal{B}] = \sum_i \sum_{\theta_i} \rho_i(\theta_i) U_i(\theta_i; \theta_i, \mathcal{B})$.

Reminiscent of equation (3) in Section 2, we obtain the belief-management representation of the remaining upward incentive constraints. For any i and $\theta_i' > \theta_i$

$$\sum_{k=\theta_i}^{\theta_i'-1} \frac{\rho_i(k)}{p(k)} \left(U_i(k; k, \mathcal{B}) - U_i(k; k{+}1, \mathcal{B})\right) - \frac{\rho(\theta')}{p(\theta')} \left(U(\theta; \theta', \mathcal{B}) - U(\theta'; \theta', \mathcal{B})\right) \geq 0,$$

which together with condition (GI) from Proposition 5 form the set of potential constraints, C_F. Next, we state a problem analogous to the dual problem in Section 2,

$$\max_{\mathcal{B} \in \{\mathcal{B}\}_p} \; \mathbb{E}[\Psi|\mathcal{B}] + \mathbb{E}[U|\mathcal{B}] \quad s.t. \; C_F. \tag{$\mathrm{P}^{\mathcal{B}}_{max}$}$$

Proposition 6 (Duality)**.** *Suppose Assumption 1 to 3 hold and fix the set of signal realizations to a singleton. A mechanism solves* (P_{min}) *if and only if its reduced form,* $(z, \gamma) = M(\mathcal{B}^*)$*, and* $\mathcal{B}^*$ *solves* ($\mathrm{P}^{\mathcal{B}}_{max}$)*.*

Proposition 6 follows from rearranging first-order conditions. We summarize the argument by providing a heuristic derivation for the case $\gamma(1,1) \geq \gamma(k,n)$, that is, a match $(1,1)$ is most likely to escalate. The escalation rule at the optimum is $\gamma(1,1)\overline{g}_{\mathcal{B}}$, and $\overline{g}_{\mathcal{B}}$ is the worst escalation rule for $\mathcal{B}$. Using Lemma 3 we rewrite

$$z_i(\theta_i) = V_i(1, (p, \underline{\rho}^V)) + \sum_{k=2}^{\theta_i} y_i(k{-}1; k) - \sum_{k=1}^{\theta_i} y_i(k; k).$$

At the optimum the probability of settlement corresponds to the probability-weighted sum of settlement values. Thus, substituting for y_i and forming expectations we get

$$\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) = \sum_i V_i(1, (p, \underline{\rho}^V)) - \gamma(1,1)\mathcal{Q}(\mathcal{B}) \overset{!}{=} 1 - Pr(\mathcal{E}) = 1 - \gamma(1,1) R(\mathcal{B}) \tag{7}$$

which implicitly determines $\mathcal{Q}(\mathcal{B})$, $R(\mathcal{B})$ and $\gamma(1,1)$. Substituting for $\gamma(1,1)$ results in

$$Pr(\mathcal{E}) = \Big(\sum_i V_i(1,(p,\underline{\rho}^V)) - 1\Big)\frac{R(\mathcal{B})}{\mathcal{Q}(\mathcal{B}) - R(\mathcal{B})} = \Big(\sum_i V_i(1,(p,\underline{\rho}^V)) - 1\Big)\frac{1}{\frac{\mathcal{Q}(\mathcal{B})}{R(\mathcal{B})} - 1}.$$

Any $\mathcal{B}$ that solves (P_{min}), solves

$$\max \frac{\gamma(1,1)\mathcal{Q}(\mathcal{B})}{\gamma(1,1)R(\mathcal{B})}.$$

Substitute for $\gamma(1,1)\mathcal{Q}(\mathcal{B})$ and for $\gamma(1,1)R(\mathcal{B})$ via (7). Bayes' rule implies $\gamma_i(\theta_i)/Pr(\mathcal{E}) = \rho_i(\theta_i)/p(\theta_i)$, and substituting for z_i and y_i yields ($\text{P}^{\mathcal{B}}_{max}$).

Finally, we allow the arbitrator to reveal additional information to players. She does so via the signaling function, Σ. As in the literature on Bayesian persuasion, such revelation can improve the outcome by linearizing convexities in the arbitrator's objective. Yet, the concavification techniques from that literature do not apply directly.[18] Instead we replace the escalation game by its random version $\widehat{\mathcal{E}}$.

Intuitively, $\widehat{\mathcal{E}}$ augments $\mathcal{E}$ by an omniscient designer who announces the realization σ of the random variable Σ. Thereafter players update beliefs to $\mathcal{B}(\sigma)$ and play $\mathcal{E}$ under these beliefs. Any realization $\mathcal{B}(\sigma)$ implies the familiar continuation payoffs $U_i(m_i;\theta_i,\mathcal{B}(\sigma))$.

The reporting strategy within the mechanism is a function of the expectations over possible information structures. Thus, the expected continuation payoff is $\widehat{U}_i(m_i;\theta_i,\mathcal{B}(\Sigma)) = \sum_\sigma Pr(\sigma)U_i(m_i;\theta_i,\mathcal{B}(\sigma))$. Using $\widehat{U}$, it is straightforward to obtain $\mathbb{E}[\widehat{\Psi}_i|\mathcal{B}(\Sigma)]$, $\mathbb{E}[\widehat{U}_i|\mathcal{B}(\Sigma)]$, $\widehat{C}_F$, and the generalization of ($\text{P}^{\mathcal{B}}_{max}$)

$$\max_{\mathcal{B}(\Sigma)} \mathbb{E}[\widehat{\Psi}_i|\mathcal{B}(\sigma)] + \mathbb{E}[\widehat{U}_i|\mathcal{B}(\sigma)] \quad s.t.\ \widehat{C}_F. \tag{$\text{P}^{\mathcal{B}(\Sigma)}_{max}$}$$

Theorem 2 (Duality of problems)**.** *Suppose Assumption 1 to 3 hold. A mechanism solves* (P_{min}) *if and only if its reduced form* $(z,\gamma) = M(\mathcal{B}^*(\Sigma))$ *and* $\mathcal{B}^*(\Sigma)$ *solves* ($\text{P}^{\mathcal{B}(\Sigma)}_{max}$)*.*

Theorem 2 displays a general characterization of the economic forces. It shows that the intuition from Section 2 holds for a large class of escalation games. Problem ($\text{P}^{\mathcal{B}(\Sigma)}_{max}$) has an analogue in optimal auction design (Myerson, 1981). The main difference is that—although types are ordered due to Assumption 2 and 3—the term $\Psi_i(\theta_i,\mathcal{B}) + U_i(\theta_i;\theta_i,\mathcal{B})$ is non-linear in the arbitrator's choice. These non-linearities complicate further characterization extremely. As we have seen in Section 2, already for the binary case results differ drastically between certain game forms.

In game forms such as the litigation contest in the ADR problem, players' behavior is very sensitive to information. Hence *both* the welfare channel and the discrimination channel are important. Moreover, in these game forms players also gain a large benefit from an information advantage. This follows because action choices depend mainly on

[18]These techniques rely on the assumption of type-independent beliefs, however, $\mathcal{B}$ specifies type-dependent beliefs. In the supplementary appendix G we show that, equivalently, we can concavify the supporting Lagrangian function.

beliefs rather than types. As a consequence, most important for the success of arbitration is which *action distributions* a belief system induces.

In contrast, the constant-sum nature of the model of Hörner, Morelli, and Squintani (2015) implies that the welfare channel becomes irrelevant. Moreover, its ex-post nature implies that the ability premium is constant. As a consequence, most important for the success of arbitration is which *type distributions* a belief system induces.

Which of the two channels dominates depends on the details of the escalation game. However, some features are common regardless of that game. Finding the optimal information structure *conditional on failed arbitration* is of first-order importance to the arbitrator. That information structure balances discrimination and efficiency conditional on escalation. It is the solution to problem ($\mathrm{P}^{\mathcal{B}(\Sigma)}_{max}$) which offers a pure information-design formulation. Moreover, reducing inefficiencies after escalation is desired although the arbitrator is agnostic about outcomes after escalation.

In the remainder of this section we consider further properties that are useful when characterizing the optimal solution for a given escalation game.

The Remaining Constraints. Similar to most reduced-form approaches in the literature the role of the general implementability constraints, (GI), is hard to interpret. We address it in Section 5. Extending the model by allowing direct utility transfers makes (GI) redundant. Alternatively, we recommend to follow the literature on reduced-form auctions using the guess and verify approach. First, we ignore (GI) and compute the (unconstrained) optimum. If (GI) fails, we include them and reoptimize.

More interesting is the role of upward incentive constraints. In principle both local and global upward incentive constraints can be relevant which makes it potentially difficult to make meaningful statements about the impact of these constraints. In the remainder of this section, we state several results and conditions that help overcoming this additional tractability issue. We begin with a simple corollary to Theorem 2 regarding the connection between these constraints and the signaling function Σ.

Corollary 2. *Consider the solution to* ($\mathrm{P}^{\mathcal{B}}_{max}$)*, ignoring C_F. The following holds*

- *if that solution satisfies C_F, the optimal signal structure is degenerate, and*
- *if for each violated constraint, the corresponding constraint of the other player holds, the symmetrizing signal from Section 2 is optimal and sufficient to satisfy C_F.*

Corollary 2 holds for the same reasons that Proposition 2 holds in the ADR case. Given any information structure, signals improve by linearizing convexities. If the unconstrained solution to ($\mathrm{P}^{\mathcal{B}}_{max}$) satisfies the constraints in C_F, there is no room for such linearizations. Signals are superfluous and the arbitrator implements an information structure with probability 1. Symmetrizing overcomes potential "holes" the constraints impose on $\{\mathcal{B}\}_p$ around symmetric information structures.

If Corollary 2 does not apply, solving the unconstrained problem ($\mathrm{P}^{\mathcal{B}}_{max}$) remains a helpful first step. The solution determines the ability premium $D_i(m_i; \theta_i, \mathcal{B})$ and the conditional likelihood that type m_i occurs, $\rho_i(\theta_i)$, via equation (6). The following proposition provides a sufficient condition assuring that the solution is a candidate mechanism net of local upward incentive constraints only.

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

$$\frac{\rho_i(m_i)}{p_i(m_i)} D_i(m_i;\theta_i,\mathcal{B}) \quad \textit{is non-increasing in } m_i, \tag{8}$$

Given Proposition 7 we can use a simple algorithm to obtain the solution. We first solve the unconstrained problem ($\mathrm{P}^{\mathcal{B}}_{max}$). If some local upward incentive constraints bind, we use that constraint to replace one belief in $\mathcal{B}$ and solve over the constraint set. We do so until we have found the optimum. Since each type has either one or two binding constraints, we are guaranteed to find the solution.

Proposition 7 requires a monotone solution to the unconstrained problem. While such a result appears intuitive, it cannot be guaranteed for all games. We conclude this section by providing two sufficient conditions on escalation games that guarantee condition (8). Both conditions are obtained on a class of *type-separable escalation games.* The payoff function of these games 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 $\zeta > 0$ increasing and continuous, ϕ, c positive, twice differentiable in both arguments and strictly increasing in a_i. Moreover, we assume that ϕ is decreasing in a_{-i} and u is concave in a_i.

Constant Difference Ratio. A first approach towards ensuring (8) is to derive conditions such the optimum is monotone. A sufficient condition is that the difference ratio $D(m_i;\theta_i,\mathcal{B})/D(m_i+1;\theta_i,\mathcal{B})$ is constant in θ_i.

Proposition 8. *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,* $\mathbb{E}[c(a_i(m_i;\theta_i,\mathcal{B}),a_{-i})|m_i,\mathcal{B}] = h(m_i;\mathcal{B})g(\theta_i)$.

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

Non-Constant Difference Ratios. Even if the difference ratio is non-constant we can specify sufficient conditions. The main ingredients to the model to guarantee a monotone solution is that actions are strategic complements. If, in addition, 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 is non-decreasing, then a simple algorithm close to the one described above yields the optimal solution even absent a monotone difference ratio, that is, even in cases in which part (ii) of Definition 5 does not hold.

A simple parameterization of such an escalation game is a compact action space $a_i \in [0,1]$, $\phi(a_i, a_{-i}) = 1/2(1 + a_i - a_{-i})$, and $c(a_i, a_{-i}) = (a_i)^2 + a_i(1 - a_{-i})/(2K)$. In the supplementary appendix K we sketch the solution algorithm for models with strategic complementarities.

5 Extensions

In this section we discuss robustness and extend our analysis to other possible modes of conflict management.

Private Value of Winning. In our main model we assume that players have private information about their cost function in the escalation game. An alternative approach is that players are privately informed about their value of winning. The analysis above remains identical under this alternative formulation. We show the isomorphism between the two models formally in the supplementary appendix J.

If the arbitrator can, in addition, impose direct utility transfers, the two models differ. In the private cost model allowing for transfers leaves the economic analysis unchanged. The only change is that the general implementation constraint, (GI), becomes redundant.

Allowing for transfers in the private valuation model, in turn, allows the arbitrator to impose the (expected) cost of escalation directly as transfers within the mechanism. Thus, escalation can always be avoided. To us, such behavior seems neither realistic nor theoretically appealing. We prefer the cost interpretation.

The Irregular Case. Assumption 3 implies that incentive compatibility holds locally in one direction and global deviations are non-profitable. The problem's properties thus closely match those of standard, monotone mechanism-design problems. Applying the guess-and-verify approach common in mechanism design, we can without loss relax Assumption 3 such that (MDR) is required only to hold *at the optimum* and verify whether Proposition 7 is applicable. Arguably the assumption remains strong nonetheless.

The assumption is demanding because we use games as outside option. In such a model the single-crossing property often fails. Assumption 3 recovers some implications of single crossing and ensures tractability. An alternative, perhaps closer, way to restore these implications is to assume *linearity in types*.

Definition 7 (Linearity in Types)**.** Let G^* be player θ_{-i}'s equilibrium distribution of actions. Then, the escalation game is linear in types if there is a pair of functions n, t such that for any m_i, θ_i, and $\mathcal{B}$

$$\sup_{a_i \in A_i^{\mathcal{E}}} \sum_{\theta_{-i}} \beta_i(\theta_{-i}|m_i) \int_{A_{-i}^{\mathcal{E}}} u(\theta_i, \theta_{-i}, a_i, a_{-i}) \mathrm{d}G^*(a_{-i}|\theta_{-i}, \mathcal{B}) = n(m; \mathcal{B})\theta_i + t(m; \mathcal{B}).$$

For a game that is linear in types, the insights of Section 4 remain and the content of Theorem 2 changes only slightly. Depending on the binding constraints, the ability premium for a particular type may be either defined upwards *or* downwards but never both. The optimal mechanism aims at increasing discrimination between a type and the *next worst* type. All other arguments prevail.

Non-Convex Veto-Values. Assumption 1 imposes several properties on the veto game. Apart from property (CONV), which states that the value of vetoing is on the convex closure at the prior, the properties are either common in the literature ((HC) and (S)) or serve a well-defined purpose (OST). Property (CONV) is special to veto-constrained mechanism-design problems. In an otherwise different problem Celik and Peters (2011) provide an example where (CONV) fails and, as a result, some types reject the optimal mechanism on the equilibrium path. We avoid such failure of the revelation principle assuming (CONV) to hold at the prior. In most of the games considered in the literature, and in our ADR model, too, (CONV) is satisfied *for all* possible priors.

However, we can eliminate assumption (CONV) completely if players have access to an independent signaling device that realizes independently of their participation decisions. If players report to the signaling device before the participation decision, but its realization can be postponed until after the acceptance decision, (CONV) is redundant. In Balzer and Schneider (2018) we study such a signaling device in detail.

Enlarging the Arbitrator's Signal Space. We restrict the arbitrator's signal space to public messages to focus on the duality of the problems. In the dual problem, the arbitrator has to solve an information-design problem for example via the methods from Mathevet, Perego, and Taneva (2017). The dual remains under a larger signal space, and results change neither conceptually nor qualitatively.

Limited Commitment by the Players. The main focus of our analysis is to highlight how the greater strategic environment around a mechanism shapes the mechanism itself. To focus on this channel we otherwise aim at staying as close as possible to standard mechanism-design problems. In particular, we assume that players have full-commitment power once the mechanism is accepted. A second important type of conflict management known as "mediation" allows players to invoke litigation at any point during the process rather than only at the beginning.

The only difference to mediation is participation constraints hold ex post. In the setting of Hörner, Morelli, and Squintani (2015) optimal mediation can achieve the same result as optimal arbitration. Their techniques do not generalize to all escalation games. We propose an alternative, yet similar, extension. Suppose instead of offering the settlement shares publicly, the designer can *privately* offer each party their share. On-path escalation is triggered via an unacceptable settlement offer and a private recommendation to reject it. Hence accepting an offer does not imply settlement. The opponent may still reject.

Using this procedure, the designer creates private information on her own. She can exploit this private information by initiating "seemingly unnecessary escalations" with a small probability on the equilibrium path. That is, breakdown occurs with positive probability regardless of players' reporting behavior. As the probability of "seemingly unnecessary escalations" approaches 0, the value of the objective converges to that without such escalations. The designer reacts to *any* deviation by implementing the worst possible belief system of the deviating type.[19] In the ADR model such a procedure is sufficient to guarantee a settlement rate arbitrarily close to the arbitration result.

Maximizing Players' Welfare. We choose the objective such that the arbitrator's preferences are as simple as possible. She focuses exclusively on achieving settlement. In principle, the arbitrator may have preferences about the outcome of the escalation game, too. She may be willing to give up some settlement solutions to decrease overall inefficiency in the continuation game.

Theorem 1 remains as it is under that objective, and adjusting ($\mathrm{P}^{\mathcal{B}}_{max}$) is straightforward. Suppose the arbitrator cares about efficiency in the escalation game. Her objective now assigns more weight on the second term, $\mathbb{E}[U_i|\mathcal{B}]$, of the objective. In the setting of Section 2 this change leads to *larger asymmetries* in the continuation game, while type-independence prevails. More generally, the derivative of the objective of ($\mathrm{P}^{\mathcal{B}}_{max}$) changes to

$$\frac{\partial \mathbb{E}[\Psi|\mathcal{B}]}{\partial \mathcal{B}} + \frac{\partial \mathbb{E}[U|\mathcal{B}]}{\partial \mathcal{B}} h_U(\mathcal{B}), \qquad \text{with } h_U \geq 1,$$

if the arbitrator also cares about maximizing joint surplus.

Implementing Reduced-Form Mechanisms. Our main results are on reduced-form mechanisms. The general implementation conditions, (GI), determine whether the solution is implementable. They are redundant in the model of Section 2 and in those from the literature. In principle, however, they are an additional constraint to the problem. Two extensions to our model are particularly related to these conditions.

The first considers correlated types. If types are correlated, the arbitrator exploits correlation via the settlement value in the same way as in Crémer and McLean (1988). She offers a "side bet" over the opponent's types to relax incentive compatibility. To achieve full efficiency unlimited utility transfers are necessary. Thus, in our setup first-best is not implementable even with correlated types. All ex-post settlement outcomes split the pie, a restriction governed by (GI). Naturally, the conditions bind when types are correlated. Otherwise, the logic of Crémer and McLean (1988) applies.

The second extension concerns additional transferable utility. If the arbitrator could impose (binding) utility transfers to the settlement allocation she is no longer restricted by the general implementability condition, (GI). Then, any reduced-form mechanism is implementable ex-post with the choice of appropriate transfers. However, such a transfer rule may include that players' ex-post payoff is negative. They might lose the good *and* pay a transfer to their opponent.[20]

[19] Correia-da-Silva (2017) and Gerardi and Myerson (2007) use similar techniques in another context.

[20] Naturally, combining the two extensions leads to full settlement. However, in real world scenarios such a combination is at most a rough benchmark.

6 Conclusion

We provide a general approach to optimal conflict management within a strategic environment partially beyond the designer's control. We propose an economically intuitive dual problem to the mechanism-design problem that links properties of the escalation game to the optimal mechanism. The dual highlights that optimal conflict management is driven by how information release during conflict management affects action choices under escalation. We reduce the main objective of optimal conflict management to the choice of an information structure conditional on escalation.

We apply our general result to optimal alternative dispute resolution in the shadow of a court. We show that optimal ADR ensures that a disputant's behavior within ADR has no effect on the information she obtains about her opponent. Applying this rule is the only way to ensure that a deviator gains no information advantage by deviating. In addition, optimal ADR guarantees an asymmetric information structure if no settlement solution results. That way, litigation after a failed settlement attempt is less inefficient than if ADR was not available. Disputant's incentives to participate are stronger.

Both results are absent in existing models in the literature which restricts attention to escalation games with ex-post equilibrium neglecting the behavioral externality. We include this option, and our belief-management approach provides an economically intuitive objective function exclusively based on the escalation game. Thereby, we not only increase tractability but also reveal the intuition behind optimal conflict management for a given escalation game.

Our results suggest a number of directions for future research. In this paper we focus on conflict management that aims to avoid a costly resolution of a dispute. However, the property that mechanisms interact with a greater strategic environment is relevant beyond conflict management. In fact, in many real-world problems a designer only controls part of the strategic environment while the information obtained through the mechanism is relevant beyond the mechanism. Problems along this line include antitrust measures, financial regulation, and international treaties of any sort.

Similar to our discussion here, information obtained during negotiations becomes valuable in future interactions, making the continuation game information sensitive. This effect, in turn, influences the design of institutions. Although many details may differ, our results suggest that there is a connection between the mechanism-design problem in restricted environments and the information-design problem in the mechanism's surroundings. Extending our results in that direction is natural, but it is beyond the scope of this paper. We leave it to future research.

Appendix

Organization: Appendix A proves Theorem 1. Appendix B proves Theorem 2. Appendix C contains further details on the ADR model from Section 2.

A Belief Management and Proof of Theorem 1

A.1 Proof of Proposition 4

Proof. Full settlement implies pooling. The sum of players' expected payoffs is constant in their types and can be set to 1. Given p, $V_i(\theta_i, (p, \rho^V))$ decreases in θ_i. If $\sum_i V_i(1, (p, \rho^V)) > 1$ for all ρ^V, then all pooling solutions violate at least one player's participation constraints. Conversely, the pooling solution $x_i = V_i(1, (p, \rho^V))$ implements full settlement. By symmetry a pooling solution exists if and only if an equal-split pooling-solution exists. □

A.2 Arbitration Mechanisms and Belief Management

We prove Theorem 1 in steps. We include potential public signals in our formal argument. A realization σ occurs with probability $Pr(\sigma)$. The realization of random variable $\mathcal{B}(\Sigma)$ is $\mathcal{B}(\sigma)$. Step 0 extends definitions from the main text including public signals. Steps 1–4 correspond to those observations in the main text indexed by the same numbers.

Step 0: Extending Definitions. The signaling function Σ determines a joint probability of escalation and the realization of σ, $\gamma^\sigma(\theta_i, \theta_{-i})$, as a function of players' reports. Individual beliefs, $\beta_i(\theta_{-i}|m_i, \sigma)$, are probability distributions conditional on own reports and σ. Other expressions extend in the natural way, $\gamma_i^\sigma(\theta_i) := \sum_{\theta_{-i}} p(\theta_{-i})\gamma^\sigma(\theta_i, \theta_{-i})$ and $y_i^\sigma(m_i; \theta_i) := \gamma^\sigma(m_i)U_i(m_i; \theta_i, \mathcal{B}(\sigma))$. Players' commitment power to accept settlement solutions implies that realizations average out in $z_i(\theta_i)$. Absent additional signals all expressions collapse to those in the main text.

Definition 8. An escalation rule γ^σ implements $\mathcal{B}(\sigma)$ if $\mathcal{B}(\sigma)$ is consistent with Bayes' rule under γ^σ.

Step 1: Homogeneity. We show $\mathcal{B}(\sigma)$ is homogeneous of degree 0 w.r.t. γ^σ via the following claim.

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

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

If $\mathcal{B}(\sigma)$ is homogeneous w.r.t. γ^σ so is U_i; $\gamma_i^\sigma(\theta_i)$ is homogeneous of degree 1 by definition and so is $y_i^\sigma(m_i; \theta_i)$.

Step 2: Worst Escalation Rule. We show that $\mathcal{B}(\Sigma)$ determines $Pr(\mathcal{G})$ up to $|\Sigma|$ real numbers. That is, the set of all escalation rules implementing a given lottery, $\mathcal{B}(\Sigma)$, is defined up to the real numbers $\{\alpha^\sigma\}_\sigma$. The escalation probability is linear in any α^σ. If the lottery is degenerate, then the worst-escalation rule is uniquely defined.

Fix a random consistent belief systems $\mathcal{B}(\Sigma)$. For each $\mathcal{B}(\sigma)$ take *some* escalation rule $\hat{\gamma}^\sigma$ that implements the belief system. Step 1 implies that the set of escalation rules implementing $\mathcal{B}(\Sigma)$ satisfies $Pr(\mathcal{G}) = \sum_{(\theta_A,\theta_B)} p(\theta_A)p(\theta_B)\Big(\sum_\sigma \alpha^\sigma \hat{\gamma}^\sigma(\theta_A,\theta_B)\Big)$. Let $\mathcal{A}$ be the set of all $\{\alpha^\sigma\}_\sigma$, with α^σ such that $\forall(\theta_A,\theta_B)$, $\alpha^\sigma\hat{\gamma}^\sigma(\theta_A,\theta_B) \leq 1$ and $\hat{\gamma}(\theta_A,\theta_B) = \sum_\sigma \alpha^\sigma\hat{\gamma}^\sigma(\theta_A,\theta_B) \leq 1$. $\mathcal{A}$ determines all escalation rules implementing $\mathcal{B}(\Sigma)$. If $\mathcal{B}(\Sigma)$ is a singleton, the largest element of $\mathcal{A}$ determines the worst escalation rule uniquely.

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

Claim. For any θ_i with $z_i(\theta_i) > 0$ either the participation 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 either $z(\theta_i) = 0$ or one of the above constraints is satisfied with equality, and all constraints remain satisfied. □

If $\Theta_i^{IC} \subset \Theta$ is the set of types with some binding incentive constraint and Θ_i^{PC}, Θ_i^0 its analogues for participation and non-negativity constraints, then $\Theta_i^{PC} \cup \Theta_i^{IC} \cup \Theta_i^0 = \Theta$. In addition, let $\Theta_i^I(\theta_i) \subset \Theta$ be the set of types such that θ_i's incentive constraints regarding mimicking any of these types holds 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 $\theta_i \notin \Theta_i^{PC} \cup \Theta_i^0$ 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 $\Theta_i^I(\theta_i) \cap \{\Theta_i^{PC} \cup \Theta_i^0 \cup \{\Theta_i^I(k)\}_{k \notin \widehat{\Theta}_i}\} = \emptyset$. □

Claim. z_i is linear in α^σ given $\mathcal{B}(\Sigma)$.

Proof. If $\theta_i \in \Theta_i^0$, z_i is constant and thus linear in α^σ. Now consider $\theta_i \in \Theta_i^{PC}$. Then, $z_i(\theta_i) = V_i(\theta_i,(p,\rho^V)) - y_i(\theta_i;\theta_i)$. The first term of the RHS is a constant, the second is linear in α^σ since $y_i(m_i;\theta_{-i}) = \sum_{\sigma\in\Sigma} y_i^\sigma(m_i;\theta_i)$ which is linear by step 1. Finally, for any $\theta_i \in \Theta_i^{IC}$, the incentive constraint is $z_i(\theta_i) = z_i(\theta_i') + y_i(\theta_i';\theta_i) - y_i(\theta_i;\theta_i)$ for any $\theta_i' \in \Theta_i^I(\theta_i)$. Given $z_i(\theta_i')$, linearity holds because y_i is linear in α^σ by step 1. Now, either $\theta_i' \in \Theta_i^{PC} \cup \Theta_i^0$, 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 $\{\Theta_i^p\}_p$ such that for every $\theta_i \in \Theta_i^p$, $\Theta^I(\theta_i) \in \Theta_i^p$. Again using step 3a, $\exists\ \theta_i \in \Theta_i^p$ s.t. $\theta_i \in \Theta_i^{PC} \cup \Theta_i^0$ and each z_i is entirely determined by additively separable, linear elements $y_i(\cdot;\cdot)$ and $V_i(\cdot,(\cdot,\cdot))$. V_i is independent of σ and each y_i is a weighted sum of all y_i^σ. Substituting into the expected settlement share and collecting terms, we can find a set of functions $H_i(\gamma^\sigma)$ solving

$$\begin{aligned}&\sum_{\theta_i} p(\theta_i) z_i(\theta_i) = \\ &-\sum_\sigma H_i(\gamma^\sigma) + \sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i) V_i(\theta_i,(p,\rho^V)) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i) \sum_{k \in \Theta^I(\theta_i)} V_i(k,(p,\rho^V))\end{aligned} \tag{9}$$

Let $H_i(\{\gamma^\sigma\}_\sigma) := \sum_\sigma H_i(\gamma^\sigma)$. Further let $P_i(\Theta_i^0) := \sum_{\theta_i \in \Theta_i^0} p(\theta_i)$. Straightforward algebra implies $H_i(\{\alpha^\sigma\gamma^\sigma\}_\sigma) = \sum_\sigma \Big(P_i(\Theta_i^0)(\alpha^\sigma - 1) H_i(\gamma^\sigma) + H_i(\gamma^\sigma)\Big)$. Thus, $H_i(\{\alpha^\sigma\gamma^\sigma\}_\sigma)$ is linearly increasing in α^σ given γ^σ.

Step 4: Determining $\{\alpha^\sigma\}_\sigma$ via resource constraint. An arbitration outcome is only feasible if the ex-ante expected settlement values are weakly lower than the probability of settlement, (5). That is, $\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \le 1 - Pr(\mathcal{G})$, where the RHS is strictly lower than 1 by Assumption 1. By step 1 any escalation rule $\{\alpha^\sigma\gamma^\sigma\}_\sigma$ implements the same $\mathcal{B}(\Sigma)$. If each $\alpha^\sigma\gamma^\sigma$ is feasible then $\{\alpha^\sigma\gamma^\sigma\}_\sigma$ satisfies (5). By step 3b we can rewrite (5) as

$$\sum_i v_i(V_i(\Theta,(p,\rho^V))) - 1 \le \sum_\sigma \left(\sum_i \Big(P_i(\Theta_i^0)(\alpha^\sigma - 1) H_i(\gamma^\sigma) + H_i(\gamma^\sigma)\Big) - Pr(\mathcal{E},\sigma)\right) \tag{5'}$$

where $v_i(V_i(\Theta,(p,\rho^V))) := \sum_i \sum_{\theta_i \in \Theta_i^{IC}} \sum_{k \in \Theta_i^I(\theta_i)} p(k) \left[\mathbb{1}_{PC}(\theta_i) V_i(\theta_i,(p,\rho^V))\right]$ is a probability weighted sum of veto values for types with binding participation constraint. Given $\Theta_i^{PC}, \Theta_i^{IC}, \Theta_i^0$, and $\{\Theta_i^I(\theta_i)\}_{\theta_i}$ the LHS is independent of the designer's choice.

Let $\{\alpha^\sigma\gamma^\sigma\}_\sigma$ implement $\mathcal{B}(\Sigma)$, then we can write the RHS as

$$\underbrace{\sum_\sigma \left(\sum_i \Big(P_i(\Theta_i^0)(\alpha^\sigma - 1) H_i(\gamma^\sigma) + H_i(\gamma^\sigma)\Big) - \alpha^\sigma \sum_{(\theta_A,\theta_B) \in \Theta} p(\theta_A) p(\theta_B) \gamma^\sigma(\theta_A,\theta_B)\right)}_{=:h(\{\alpha^\sigma\gamma^\sigma\}_\sigma)}.$$

Moreover, using the definition of H_i it follows that h is linear in α since $y_i(\theta_i;\theta_i)$ and $Pr(\mathcal{E})$ are homogeneous in $\{\gamma^\sigma\}_\sigma$. In particular, $h(\sum_\sigma \alpha^\sigma\gamma^\sigma)$ converges to a weakly positive number if every α^σ is sufficiently small. Observe that $\sum_\sigma \alpha^\sigma\gamma^\sigma \to 0$ is the full settlement solution. Thus, Assumption 1 implies that the LHS of (5') is strictly positive. In turn, $h(\{\alpha^\sigma\gamma^\sigma\}_\sigma) > 0$ because $\{\alpha^\sigma\gamma^\sigma\}_\sigma$ is part of an implementable mechanism. Therefore, the optimal $\{\alpha^\sigma\}_\sigma$ equates LHS and RHS. Thus, for any $\mathcal{B}(\Sigma)$ the minimal $Pr(\mathcal{G})$ uses an $\{a^\sigma\}_\sigma$ at the boundary of $\mathcal{A}$.

B Optimal Arbitration and Proof of Theorem 2

We construct a solution algorithm to solve for $\mathcal{M}$. We use it to prove Theorem 2.

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

Lemma 11 in the supplementary appendix I provides the corresponding formal argument.

B.1 Proof of Proposition 5

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

B.2 Proof of Lemma 3

Proof. The MDR property implies that local downward incentive compatibility is sufficient for global downward incentive compatibility.[21] Now, assume there is a type θ_i for which both incentive constraints are redundant. By the argument of step 3a on page 36 it follows that $\theta_i \in \Theta_i^{PC}$, or, $\theta_i \in \Theta_i^0$.

Assumption 1 implies that the set of types with binding participation constraints is non-empty. Otherwise full settlement is feasible. If there is exactly one type of one player with a binding participation constraint, the designer can offer an alternative mechanism: The mechanism determines at random who is assigned the role of player i and who that of $-i$ *after* players have submitted their report. Each of the two realizations satisfies the constraints. The combination satisfies constraints by symmetry. Under the alternative mechanism, no participation constraint is binding. A contradiction.

To see that exactly 1's participation constraint is binding and that $z_i(\theta_i) > 0$, consider the designer's resource constraint. Focus on the formulation (5') in the proof of Theorem 1, step 4. Assume by contradiction that the set of types with non-binding incentive constraints $\Theta_i^{PC} \cup \Theta^0 \neq \{1\}$. An upper bound on LHS of (5') is

$$\sum_i \sum_{\substack{\theta_i \in \\ \Theta_i^{PC} \cup \Theta_i^0}} \sum_{\substack{k \in \\ \Theta_i^I(\theta_i)}} p(k) \left[\mathbb{1}_{PC}(\theta_i) V_i(\theta_i, (p, \rho^V)) \right] - 1 \leq \sum_i \sum_{\substack{\theta_i \in \\ \Theta_i^{PC} \cup \Theta_i^0}} \sum_{\substack{k \in \\ \Theta_i^I(\theta_i)}} p(k) V_i(\theta_i, (p, \rho^V)) - 1.$$

Assumption 3, part (i), implies a negative upper bound if $\Theta_i^{PC} \neq \{1\}$ contradicting Assumption 1. Given the set of types with binding participation constraint, $\Theta_i^{PC} = \{1\}$, and $\Theta_I^0 = \emptyset$, (MDR) together with step 3a on page 36 implies that it is without loss to assume binding local downward incentive compatibility. □

B.3 The Lagrangian Problem

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

Lemma 4. *The Lagrangian approach yields the global optimum.*

Proof. We use theorem 1 in Luenberger (1969) to show that the Lagrangian approach is sufficient. Let T be the set of Lagrangian multiplier, with element t. Further, let $G(\cdot)$ be the set of inequality constraints and $Pr(\mathcal{E})$ a function from choices to escalation probabilities. Define $\mathrm{w}(t) := inf\{Pr(\mathcal{E}) | cs = (\gamma, z, \Sigma) \in CS, G(cs) \leq t\}$. The Lagrangian is sufficient for a global optimum if $\mathrm{w}(t)$ is convex.

Assume for a contradiction that $\mathrm{w}(t_0)$ is not convex at t_0. Then, there is t_1, t_2 and $\lambda \in (0,1)$ such that $\lambda t_1 + (1-\lambda) t_2 = t_0$ and $\lambda \mathrm{w}(t_1) + (1-\lambda)\mathrm{w}(t_2) < \mathrm{w}(t_0)$. For $j \in \{1, 2\}$ let

[21] The proof is along the standard argument that the monotone likelihood ratio implies sufficiency of local incentive compatibility in standard mechanism-design problems. Our version of the proof is in Section F of the supplementary material.

$cs_j = (\gamma[j], z[j], \Sigma[j])$ describe the optimal solution, such that $Pr(\mathcal{E})(cs_j) = \mathrm{w}(t_j)$. Then, consider the choice cs_0 such that $z[0] = \lambda z[1] + (1-\lambda)z[2]$, $\gamma[0] = \lambda\gamma[1] + (1-\lambda)\gamma[2]$ and $\Sigma = \{1, 2\}$, with $Pr(\sigma = 1) = \lambda$ and $\gamma^{\sigma=j} = \gamma[j]$. By construction constraints are satisfied and and the solution value equals that of the convex combination

$$\mathrm{w}(t_0) = Pr(\mathcal{E})(cs_0) = \sum_{\sigma\in\{1,2\}} Pr(\sigma)Pr(\mathcal{E}|\sigma) = \tilde{\alpha}\mathrm{w}(t_1) + (1-\tilde{\alpha})\mathrm{w}(t_1)$$

A contradiction. □

We continue under Assumption 3. The approach absent Assumption 3 is similar, yet notationally more involved. We describe it in the supplementary appendix G. First, we define the conditional type probabilities.

Definition 9 (Conditional Type Probabilities)**.** Let $\rho_i(\cdot|\sigma)$ be the probability distributions over i's types conditional on escalation and σ. It is the solution to the system of linear equations $\rho_i(\theta_i|\sigma) = \sum_{\theta_{-i}} \beta_{-i}(\theta_i|\theta_{-i}, \sigma)\rho_{-i}(\theta_{-i}|\sigma)$. The conditional probability distribution of a profile is $\rho(\theta_i, \theta_{-1}|\sigma) := \beta_{-i}(\theta_i|\theta_{-i}, \sigma)\rho_{-i}(\theta_{-i}|\sigma)$. The set of all $\rho(\theta_i, \theta_{-1}|\sigma)$ is $\boldsymbol{\rho}(\sigma)$.

Given Σ, $\mathcal{B}(\sigma)$ and $\boldsymbol{\rho}(\sigma)$ are isomorphic. We define the set of Lagrangian multipliers $\nu^i_{\theta,\theta+1}$ for incentive constraints, λ^i_θ, for the participation constraints, δ, for the designer's resource constraint, ζ^i_θ for non-negativity of z_i, and η_Q for the general implementability constraints of the reduced-form mechanism from Proposition 5. Finally, μ_{θ_A,θ_B} is the multiplier on the consistency constraint. We divide all multipliers by δ and obtain the set $\{\tilde{\nu}^i_{\theta,\theta'}, \tilde{\lambda}^i_\theta, 1, \tilde{\zeta}_{\theta^i}, \tilde{\eta}_Q, \tilde{\mu}_{\theta_A,\theta_B}\}$. Let $\tilde{e}_i(\theta) = p(\theta) \sum\limits_{Q|\theta\in Q} \eta_Q$. Furthermore, we define the aggregation up to type θ using capital letters, $\widetilde{\Lambda}^i(\theta) := \sum_k^\theta \tilde{\lambda}^i_k$, and $\widetilde{E}^i(\theta)$ and $\widetilde{Z}^i(\theta)$ in a similar way. Finally, $\mathsf{m}^{\mathsf{i}}_\theta := p(\theta) + \tilde{e}^i_\theta - \tilde{\zeta}^i_\theta$, $\mathsf{M}^{\mathsf{i}}(\theta) := \widetilde{\Lambda}^i(\theta) - \sum_{k=1}^{k=\theta} p(k) - \widetilde{E}^i(\theta) + Z^i(\theta)$. Moreover, define $\tilde{N}^i(\theta,\theta{+}1) := -(\sum_{k=1}^\theta \sum_{\hat{\theta}>\theta}^K \tilde{\nu}^i_{\theta,\hat{\theta}}) + \tilde{\nu}^i_{\theta,\theta+1}$. We state the transformed Lagrangian objective as a corollary to the more general solution discussed in the supplementary appendix G.

Corollary 3. *Suppose Assumption 1 to 3 hold. The lottery* $\{Pr(\sigma), \boldsymbol{\rho}(\sigma)\}_{\sigma\in\Sigma}$ *is an optimal solution to the designers problem if and only if* each $\boldsymbol{\rho}(\sigma)$*, maximizes*

$$\begin{aligned}\widehat{\mathcal{L}}(\mathcal{B}(\sigma)) :=& \mathcal{T}(\mathcal{B}(\sigma)) + \sum_i \Bigg[\sum_{\theta=1}^K \rho_i(\theta|\sigma)(\frac{\mathsf{m}^{\mathsf{i}}_\theta}{p(\theta)}) U_i(\theta;\theta,\mathcal{B}(\sigma)) \\ &+ \sum_{\theta=1}^{K-1} \frac{\tilde{\nu}^i_{\theta,\theta+1} + \mathsf{M}^{\mathsf{i}}(\theta) - \tilde{N}^i(\theta,\theta{+}1)}{p(\theta)} \rho_i(\theta|\sigma)\{U_i(\theta;\theta,\mathcal{B}(\sigma)) - U_i(\theta;\theta{+}1,\mathcal{B}(\sigma))\} \\ &- \sum_{\theta=2}^K \sum_{\theta'=1}^{\theta-1} \frac{\tilde{\nu}^i_{\theta,\theta'}}{p(\theta)} \rho_i(\theta|\sigma)\left\{U_i(\theta;\theta',\mathcal{B}(\sigma)) - U_i(\theta;\theta,\mathcal{B}(\sigma))\right\} \Bigg],\end{aligned} \tag{10}$$

with

$$\mathcal{T}(\mathcal{B}(\sigma)) := \sum_{Q\in Q^2}\sum_{(\theta_A,\theta_B)\in\tilde{Q}}[\rho(\theta_A|\sigma)\beta_1(\theta_B|\theta_A,\sigma)]\tilde{\eta}_Q - \sum_{\theta_A\times\theta_B}\frac{\rho_1(\theta_A|\sigma)\beta_1(\theta_B|\theta_A,\sigma)}{p(\theta_A)p(\theta_B)}\tilde{\mu}_{\theta_A,\theta_B}. \tag{11}$$

Hence $\boldsymbol{\rho} := \sum_\sigma Pr(\sigma)\boldsymbol{\rho}(\sigma)$ *is a maximizer of the concave closure of the above function. The following holds by complementary slackness*

- *The resource constraint from equation* (5) *binds, hence* $\delta > 0$.
- *If the optimal reduced-form mechanism is implementable, then the constraints from Proposition 5 are redundant and* $\widetilde{E}_i(\theta) = \tilde{e}^i_\theta = \widetilde{Z}^i(\theta) = \tilde{\zeta}^i_\theta = 0$.
- *Downward local incentive constraints bind, thus* $M_i(\theta) > 0$. *If, in addition, all upward incentive constraints are redundant, then* $\tilde{\nu}^i_{\theta,\theta'} = 0$ *for all* $\theta' \geq \theta$ *and* $\tilde{\nu}^i_{\theta,\theta+1} = \tilde{N}^i(\theta,\theta{+}1) = 0$.
- *Local downward incentive constraints are sufficient for global downward incentive constraints.*

Results follow from algebraic manipulation of the initial Lagrangian objective using Lemma 3 to identify binding constraints. Manipulations proceed alongside the discussion of Proposition 6. A full description is in Section G of the supplementary material.

B.4 Proof of Proposition 6

Proof. With access to signals the designer can implement spreads over consistent post-escalation belief systems. Then, (i) the Lagrangian approach yields the global maximum, and (ii) the optimal solution lies on the concave closure of the Lagrangian function over consistent post-escalation belief systems. Without access to signals, (i) a critical point of the Lagrangian objective is only necessary but not sufficient for global optimality, (ii) every optimal solution must be a local maximum of the Lagrangian objective (but not of its concave hull), and (iii) constraints have to hold for the ex-post realized belief system (rather than for the lottery over realized belief systems). Despite these differences, we still can use the form of the Lagrangian function stated in Corollary 3. Take the first two terms of the Lagrangian in Corollary 3 as the objective due to the binding constraints from Lemma 3, set the Border multipliers $\tilde{e}^i_\theta, \zeta^i_\theta, \widetilde{E}^i, Z^i_\theta$ to zero and add the respective constraints from Proposition 5. Consequently, $\frac{\mathsf{m}^{\mathsf{i}}_\theta}{p(\theta)} = 1$ and $\mathsf{M}^{\mathsf{i}}(\theta)/p(\theta) = w(\theta)$. The last term of (10) boils down to an expression that consists of local downward incentive constraints. The signaling term (11) is implied by consistency and the Border constraints, completing the proof. □

B.5 Proof of Theorem 2

Proof. The problem collapses to $(\mathrm{P}^{\mathcal{B}}_{max})$ if the optimal signal is degenerate and Corollary 2 applies. Suppose we are at an optimum with a non-degenerate signal. Assume that instead of the continuation game $\mathcal{E}$, an alternative continuation game $\widehat{\mathcal{E}}$ is played. $\widehat{\mathcal{E}}$ differs from $\mathcal{E}$ in that an omniscient nature first draws a realization of a state-dependent random variable Σ and communicates this to the players. Players update to $\mathcal{B}(\sigma)$ and play $\mathcal{E}$ under updated beliefs. The continuation payoff of $\hat{\mathcal{E}}$ is $\widehat{U}(m;\theta,\mathcal{B}(\Sigma))$. If $\mathcal{B}(\Sigma)$ satisfies the constraints, it is implementable. Furthermore, $\mathcal{B}(\Sigma)$ leads to a random expected ability premium, $\mathbb{E}[\widehat{\Psi}|\mathcal{B}(\Sigma)] := \sum_i \sum_\sigma Pr(\sigma)\mathbb{E}[\Psi_i|\mathcal{B}(\sigma)]$, and a random expected welfare, $\mathbb{E}[\widehat{U}|\mathcal{B}(\Sigma)] := \sum_i \sum_\sigma Pr(\sigma)\mathbb{E}[U_i(\theta;\theta,\mathcal{B}(\sigma))|\mathcal{B}(\sigma)]$. Then we proceed as in Proposition 6. □

B.6 Proof of Corollary 2

Proof. The solution to the optimization problem, $\mathcal{B}^*$, maximizes (10). By hypothesis, $\mathcal{B}^*$ is in the set of least-constrained solutions. Thus, the last two terms of equation (10) are 0. Thus, $\mathcal{B}^*$, maximizes

$$\sum_i \left[\sum_{\theta=1}^{K} \rho_i(\theta) U_i(\theta;\theta,\mathcal{B}) + \sum_{\theta=1}^{K-1} \frac{1-\sum_{k=1}^{k=\theta} p(k)}{p(\theta)} \rho_i(\theta)(U_i(\theta;\theta,\mathcal{B}) - U_i(\theta;\theta+1,\mathcal{B})) \right],$$

By construction the optimum is on the concave closure, signals do not improve. If constraints bind only one-sidedly, they do not bind under the symmetrizing signal (see also the argument in appendix B.2). □

B.7 Proof of Proposition 7

The proof follows from Lemma 6 in the supplementary appendix F.

B.8 Proof of Proposition 8

Omitted. A proof is in the supplementary appendix K.

C Additional Material to Section 2

We first provide institutional details on ADR in the US. Thereafter we give a graphical intuition behind the economics of ADR, employing the belief-management perspective. We then formally state the equilibrium objects in the litigation game. That proves Lemma 1. Thereafter we prove Proposition 1 and 2 jointly. Finally, we present additional properties of the optimal mechanism. The missing proofs of Lemma 2 and Proposition 3 are in the supplementary material.

C.1 Optimal ADR: Institutional Background

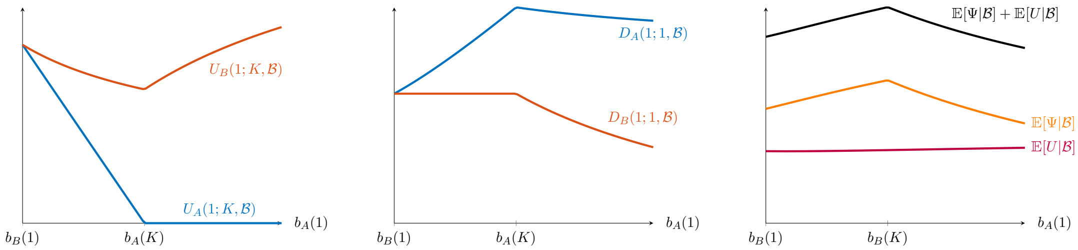
Alternative Dispute Resolution (ADR) is a tool introduced into legal systems to reduce inefficiencies through out-of-court settlements. ADR as an umbrella term describes third-party mechanisms to solve disputes other than formal litigation. Typically ADR cannot overturn the rule of law and parties return to the litigation track once ADR fails.

The court system of most developed countries is heavily overburdened. The average judge in a US district court in 2017 receives 639 newly filed cases per year. At the same time she has a stock of around 744 pending cases.[22] Most jurisdictions encourage disputants to engage in some form of ADR before starting the formal litigation process.[23]

The U.S. Alternative Dispute Resolution Act of 1998 states that courts should provide disputants with ADR-options in all civil cases. ADR is defined as "any process or

[22]There were 677 judgeships in 2017 but also 1266 vacant judgeship months. To calculate the average judge's yearly incoming cases, we use the non-vacant judgeships months.

[23]See `http://www.uscourts.gov/data-table-numbers/jci` for further data on US district courts.



(a) Deviation payoff. (b) Ability premium. (c) Objective Function.

Figure 5: *Effect of type-dependence.* $b_B(K)$ adjusts endogenously to ensure a consistent $\mathcal{B}$.

procedure, other than an adjudication by a presiding judge, in which a neutral third party participates to assist in the resolution of issues in controversy" (Alternative Dispute Resolution Act, 1998). ADR supplements the "rule of law" rather than replacing it. Ultimately, each disputant has the right to return to formal litigation.[24] Hence ADR indeed happens "in the shadow of the court."

ADR is effective with success rates substantially above 50% across time, jurisdictions, and case characteristics. In addition disputants report that ADR has an impact on the subsequent litigation process even if unsuccessful (Anderson and Pi, 2004; Genn, 1998). During the process they obtain information useful to them in litigation.

C.2 Optimal ADR: A Graphical Intuition.

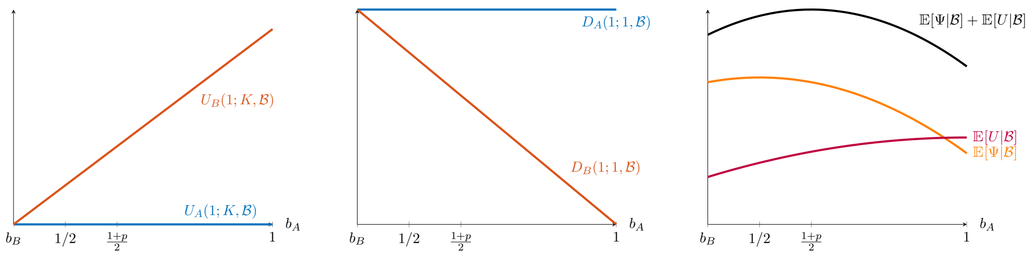
No-information trading. Consider a weak disputant after escalation. If $b_A(1) < b_A(K)$, she expects a weaker opponent than on the equilibrium path. She reacts by putting full mass on $\overline{a}_B^K$. If $b_A(1) > b_A(K)$, she expects a tougher opponent. She reacts by putting full mass (close) to 0.[25] In the first case, she shifts mass upwards to obtain a higher probability of winning. In the second case, she shifts mass downwards to cut losses.

Figure 5(a) graphs the deviation payoffs. They are minimized if $b_A(1) = b_A(K)$, the case without information advantage. Figure 5(b) depicts the ability premium. The ability premium has its maximum at $b_A(1) = b_A(K)$ because of the information advantage. By construction, the same holds for $\mathbb{E}[\Psi|\mathcal{B}]$ in Figure 5(c). However, $\mathbb{E}(U|\mathcal{B})$ changes only marginally in $b_A(1)$. The optimal mechanism eliminates the information advantage and information sets are independent of behavior in ADR.

Asymmetry. The effect of asymmetric beliefs is sketched in Figure 6. Starting at $b_A = b_B$ and gradually increasing b_A has no effect on the continuation payoff of a weak disputant A, but increases the continuation payoff of weak disputant B. The effect of asymmetry on on-path payoffs in contests is well-known. We graph it in Figure 6(c). The more asymmetric a contest, the lower the evidence level in equilibrium. Combining both effects yields a hump-shaped objective. Thus, there is an optimal degree of asymmetry.

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

[25]No maximum exists. The supremum in equation (1), $a_i = 0$, a tight upper bound.



(a) Deviation payoff. (b) Ability premium. (c) Objective Function.

Figure 6: *Effect of asymmetry.* Increasing b_A from b_B while holding beliefs type independent. The optimal solution is $b_A = (1+p)/2$ independent of the choice of $b_B \leq 1/2$.

C.3 Equilibrium in the Litigation Game and Proof of Lemma 1

We identify disputants through marginal cost θ_i. We focus on interior beliefs $b_i \in (0,1)$. The extension to the boundary cases is straightforward.

Strategies: Densities and Distributions. (cf. Siegel, 2014, for a general discussion). The distribution function $F_i^{\theta_i}(a)$ denotes the probability of θ_i choosing an action smaller than a_i. Disputant θ_i's support includes a if and only if a maximizes

$$F_{-i}(a|\theta_i) - a\theta_i = (1 - b_i(\theta_i))\, F_{-i}^K(a) + b_i(\theta_i) F_{-i}^1(a) - a\theta_i.$$

Define the partitions $I_1 = (0, \overline{a}_B^K]$, $I_2 = (\overline{a}_B^K, \overline{a}_A^K]$ and $I_3 = (\overline{a}_A^K, \overline{a}_A^1]$. We define indicator functions $\mathbb{1}_{\in I_l}$ with value 1 if $a \in I_l$ and 0 otherwise. Similar the indicator function $\mathbb{1}_{>I_l}$ takes value 1 if $a > \max I_l$ and 0 otherwise. Disputant θ_i mixes such that the opponent's first-order condition holds on the joint support. The densities are

$$f_B^1(a) = \mathbb{1}_{\in I_2} \frac{K}{b_A(K)} + \mathbb{1}_{\in I_3} \frac{1}{b_A(1)}, \qquad f_B^K(a) = \mathbb{1}_{\in I_1} \frac{K}{1 - b_A(K)},$$
$$f_A^1(a) = \mathbb{1}_{\in I_3} \frac{1}{b_B(1)}, \qquad f_A^K(a) = \mathbb{1}_{\in I_1} \frac{K}{1 - b_B(K)} + \mathbb{1}_{\in I_2} \frac{1}{1 - b_B(1)}.$$

This leads to the following cumulative distribution functions:

$$F_2^1(a) = \mathbb{1}_{\in I_2} a \frac{K}{b_A(K)} + \mathbb{1}_{\in I_3} \left(\frac{a}{b_A(1)} + F_B^1(\overline{a}_B^K) \right) + \mathbb{1}_{>I_3},$$
$$F_2^K(a) = \mathbb{1}_{\in I_1} a \frac{K}{1 - b_A(K)} + \mathbb{1}_{>I_1},$$
$$F_1^1(a) = \mathbb{1}_{\in I_3} \frac{a}{b_B(1)} + \mathbb{1}_{>I_3},$$
$$F_1^K(a) = \mathbb{1}_{\in I_1} \left(a \frac{K}{1 - b_B(K)} + F_A^K(0) \right) + \mathbb{1}_{\in I_2} \left(\frac{a}{1 - b_B(1)} + F_B^K(\overline{a}_B^K) \right) + \mathbb{1}_{>I_2}.$$

Disputants' Strategies: Interval Boundaries. The densities define the strategies up to intervals' boundaries. These boundaries are determined as follows

1. $\overline{a}_B^K$ is determined using $F_B^K(\overline{a}_B^K) = 1$, i.e., $\overline{a}_B^K f_B^K(a) = 1$ for $a \in I_1$. Substituting

yields

$$\overline{a}_B^K = \frac{1-b_A(K)}{K}.$$

2. For any $\overline{a}_A^K$, $\overline{a}_A^1$ is determined using $F_A^1(\overline{a}_A^1) = 1$, i.e., $\left(\overline{a}_A^1 - \overline{a}_A^K\right) f_A^1(a) = 1$ with $a \in I_3$. Substituting yields
$$\overline{a}_A^1 = \overline{a}_A^K + b_B(1).$$
3. $\overline{a}_A^K$ is determined by $F_B^1(\overline{a}_A^K) = 1$. That is, $\left(\overline{a}_A^K - \overline{a}_B^K\right) f_B^1(a) + \left(\overline{a}_A^1 - \overline{a}_A^K\right) f_B^1(a') = 1$ with $a \in I_2, a' \in I_3$. Substituting yields
$$\overline{a}_A^K = \overline{a}_B^K + \left(1 - \frac{b_B(1)}{b_A(1)}\right) \frac{b_A(K)}{K}.$$
4. $F_A^K(0)$ is determined by the condition $F_A^K(\overline{a}_A^K) = 1$, i.e., $F_A^K(0) = 1 - \overline{a}_B^K f_A^K(a) - \left(\overline{a}_A^K - \overline{a}_B^K\right) f_A^K(a')$ with $a \in I_1, a' \in I_2$. Substituting yields
$$F_A^K(0) = 1 - \frac{1-b_A(K)}{1-b_B(K)} - \left(1 - \frac{b_B(1)}{b_A(1)}\right) \frac{b_A(K)}{1-b_B(1)} \frac{1}{K}.$$

Change of Variables. To simplify the argument in the proof of Proposition 1 we express $\mathcal{B}$ entirely using ρ_i and $b_A(1)$. The ex-ante probability that i is strong, $\rho_i := b_{-i}(1)b_{-i}(K)/(b_{-i}(1)(1-b_i(1)) + b_i(1)b_{-i}(K))$. Given ρ_A, ρ_B we describe any $b_i(m) \in \mathcal{B}$ as a linear function of $b_A(1)$ using Bayes' rule. That is,

$$b_A(K) = \frac{\rho_B - \rho_A b_A(1)}{1-\rho_A}, \quad b_B(K) = \frac{\rho_A}{1-\rho_B}(1-b_A(1)), \quad \text{and } b_B(1) = \frac{\rho_A}{\rho_B} b_A(1).$$

Closed-form Expressions relative to $b_A(1)$. Belief fractions such as b_i/b_{-i} depend only on ρ, and we simplify further

$$\begin{aligned}
\overline{a}_B^K &= \frac{1-\rho_B-\rho_A}{K(1-\rho_A)} + \frac{\rho_A}{K(1-\rho_A)} b_A(1), \qquad & F_A(0) &= (1-b_B(m)\left(\frac{(\rho_B-\rho_A)}{(1-\rho_A)}\frac{(K-1)}{K}\right), \\
F_B(\overline{a}_B^K|m) &= (1-b_A(m)), & F_A(\overline{a}_B^K|m) &= (1-b_B(m))\left(1 - \frac{(\rho_B-\rho_A)}{(1-\rho_A)}\frac{1}{K}\right).
\end{aligned} \tag{12}$$

C.4 Proof of Proposition 1 and 2

Structure of the Proof. We prove Proposition 1 and 2 using a guess and verify approach. A constructive proof is possible but notationally intense. We omit showing that at the optimum the escalation game has a unique and monotonic equilibrium and we omit the case for small priors $p < \underline{r} := (2(K-1) - \sqrt{8-4K+K^2})/(2+3K)$. Both aspects are straightforward to verify. A more constructive version including these missing steps is in the companion paper Balzer and Schneider (2017), and a matlab program on our website calculates the optimal solution for $p < \underline{r}$. We start by guessing that the conditions in Corollary 2 hold and no additional constraint binds. In the end we verify that guess and prove Proposition 2 for the remaining cases.

Proof. ***Part A (Piecewise Linearity).*** $b_A(1) \geq b_B(1)$ implies $\rho_B > \rho_A$. Take any ρ_i that satisfies this condition. A disputant's winning probability, $F_i(\overline{a}_B^K|m)$, is linear in $b_A(1)$ since $(1-b_i(m))$ is linear in $b_A(1)$. $\overline{a}_B^K$ is linear in $b_A(1)$, too, and so are the payoffs. Finally, the virtual rent, $E[\Psi|\mathcal{B}]$ is piecewise linear with a kink at $b_A(1) = \rho_B$ and so is the objective. Multiplying with p for readability yields

$$\begin{aligned}\Xi(b_A(1)) &:= p\left(\sum_i \mathbb{E}[\Psi_i|\mathcal{B}] + \mathbb{E}[U_i|\mathcal{B}]\right) \\ &= (\rho_A+\rho_B)U_A(1;1,\mathcal{B}) - (1-p)\sum_i \rho_i U_i(1;K,\mathcal{B}) + p(1-\rho_B)U_B(K;K,\mathcal{B}).\end{aligned} \tag{13}$$

Part B (Optimality).

Step 1: Type-independence. Linearity immediately implies that the optimal $b_A(1)$ includes a point at the boundary. The relevant boundaries are the lowest value and the highest value such that the solution yields a monotonic equilibrium, and the point at which the virtual rent has a kink. These three points are

$$\underline{b} = \frac{\rho_A}{K(1-\rho_B)+\rho_B}, \qquad \overline{b} = \frac{(K-1)(1-\rho_A)+\rho_B}{K(1-\rho_A)+\rho_A}, \qquad b^* = \rho_B.$$

We guess that the optimum is at $b^* = \rho_B$ and proceed.

Step 2: Type distribution. Assuming type-independence we determine the optimal ρ_i. Using the RHS of (13) and substituting $b_A(1) = \rho_B$ yields an objective quadratic in either ρ_i. Moreover, the first-order conditions are independent of each other. The unique solution is $(\rho_A, \rho_B) = ((1-p)/2, (1+p)/2)$.[26] Second-order conditions are satisfied at the desired point and we can conclude that a local optimum exist in case we face a least-constrained problem. If $p \geq \underline{r}$, there always exists an α and thus an escalation rule such that the optimal solution satisfies the resource constraint, (5), with equality.

Step 3: Upward Incentive Constraints and Proof of Proposition 2. Downward incentive constraints hold with equality by construction. However, so far we have ignored type 1's incentive constraint, $\gamma_i(1)U_i(1;1,\widehat{\mathcal{B}}) + z_i(1) \geq \gamma_i(K)U_i(K;1,\widehat{\mathcal{B}}) + z_i(K)$. Using $z_i(K) - z_i(1) = \gamma_i(1)U_i(1;K,\widehat{\mathcal{B}}) - \gamma_i(K)U_i(K;K,\widehat{\mathcal{B}})$ the constraint becomes

$$\gamma_i(1)U_i(1;1,\widehat{\mathcal{B}}) - \gamma_i(K)U_i(K;1,\widehat{\mathcal{B}}) \geq \gamma_i(1)U_i(1;K,\widehat{\mathcal{B}}) - \gamma_i(K)U_i(K;K,\widehat{\mathcal{B}}).$$

Type independence implies $U_i(1;K,\widehat{\mathcal{B}}) = U_i(K;K,\widehat{\mathcal{B}})$. Incentive compatibility holds if

$$\gamma_i(1) \geq \gamma_i(K) \Leftrightarrow \rho_i \geq p. \tag{14}$$

This always holds for disputant A, type 1 but not for disputant B type 1 if $p > 1/3$. Now consider the following mechanism with public signals. There are two realizations, σ_A and σ_B, both equally likely. If σ_A realizes the mechanism proceeds as above, if σ_B realizes, the

[26]By continuity of the objective the same holds true if we take the objective given $b_A(1) \geq \rho_i$ instead.

mechanism flips disputants' identities. By ex-ante symmetry, the value of the problem remains constant and condition (14) holds by Assumption 1 as it becomes

$$\gamma_i^{\sigma_A}(1)+\gamma_i^{\sigma_B}(1) \geq \gamma_i^{\sigma_A}(K)+\gamma_i^{\sigma_B}(K) \Leftrightarrow \frac{1}{2} \geq p.$$

Step 4: Verifying Local Optimality. We now verify that type independence yields a local optimum. Assume to the contrary that $b_A(1) < \rho_B$ at the optimum. Substituting the claimed optimum into the objective we observe that

$$\begin{aligned} \Xi(b_A(1))|_{b_A(1)<\rho_B} = & \, F_1^K(\overline{a}_B^K)\left((1-b_A(1))\,\rho_B + p(1-\rho_B)\right) \\ & - (1-b_B(1))\left((1-p)\rho_B F_1^K(\overline{a}_B^K)\right) \\ & + \underbrace{\overline{a}_B^K}_{=(1-b_A(K))/K} \left((\rho_A{+}\rho_B)\,(K{-}1) - (1+\rho_B)Kp\right). \end{aligned} \tag{15}$$

The derivative changes sign at $b_A(1) = \rho_B$. The objective's derivative at the optimal point $\rho^* = \big(\rho_A = (1-p)/2, \rho_B = (1+p)/2\big)$ reads

$$\frac{\partial \Xi(b_A(1))}{\partial b_A(1)}|_{\rho^*} = \begin{cases} \frac{K(1-(p)^2)-(1-(p)^2)}{K(1+p)} & \text{if } b_A(1) < \rho_B \\ -\frac{K(1-(p)^2)-(1-(p)^2)}{K(1+p)} & \text{if } b_A(1) > \rho_B \\ \text{undefined} & \text{if } b_A(1) = \rho_B. \end{cases}$$

Step 5: Global Optimality. To verify global optimality plug $b_A(1) \in \{\underline{b}, \overline{b}\}$ into Ξ and observe that the outcome is inferior.

Step 6: Implementability. Plugging into (GI) verifies that the reduced form is implementable. □

C.5 Properties of the optimal mechanism

We focus on the case $p \leq 1/3$ and on solutions in the interior of the feasible set of information structures. This case leads to closed-form solutions and it is without loss to ignore the strong type's incentive constraint. Our characterization, Corollary 1, applies also to the remaining cases.[27]

An interior solution implements $b_A{=}(1{+}p)/2$ and $b_B{=}(1{-}p)/2$ via the escalation rule

$$\begin{pmatrix} \gamma_A(1,1) & \gamma_A(1,K) \\ \gamma_A(K,1) & \gamma_A(K,K) \end{pmatrix} = \alpha \begin{pmatrix} 1 & \frac{p}{1+p} \\ \frac{p(1+p)}{(1-p)^2} & \left(\frac{p}{1-p}\right)^2 \end{pmatrix},$$

where $\alpha \in (0,1]$ is a scalar. The scalar α is determined by making the arbitrator's resource constraint binding. We address it below. The probability of escalation conditional on i's

[27] The set of feasible information structures is determined by Bayes rule. Any $\mathcal{B}$ must be an element of some mean-preserving spread over the prior p (see also Kamenica and Gentzkow, 2011, for details on the procedure).

own report is $\gamma_i(m_i) := p\gamma_i(m_i, 1) + (1-p)\gamma_i(m_i, K)$. For $\alpha < 1$ we get,

$$\gamma_A(1) = \frac{2p}{1+p}\alpha, \quad \gamma_B(1) = \frac{2p}{1-p}\alpha, \quad \gamma_A(K) = 2\left(\frac{p}{1-p}\right)^2\alpha, \quad \gamma_B(K) = \frac{2p^2}{1-p^2}\alpha.$$

The arbitrator's resource constraint is always binding. The arbitrator can promise agents a settlement value only if she manages to avoid escalation. Optimality then provides the following condition

$$\sum_i (pz_i(1) + (1-p)z_i(K)) = 1 - (p\gamma_i(1) + (1-p)\gamma_i(K)). \tag{R}$$

The left-hand side is the expected sum of settlement shares, the right hand side the expected probability of settlement. Both sides are linear in α. Solving (R) for α, we obtain that α is increasing in K and decreasing in p. As $\alpha \to 1$, the optimal solution moves to the boundary.

The optimal *realized* settlement share $x_i(m_i, \theta_{-i})$ is not unique, but we can determine the *expected* settlement share, $x_i(m_i) := z_i(m_i)/(1-\gamma_i(m_i))$.

References

Anderson, H. and R. Pi (2004). *Evaluation of the Early Mediation Pilot Programs*. Tech. rep. Judical Council of California.

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 Maschler (1995). *Repeated games with incomplete information*. Cambridge, MA: MIT press.

Balzer, B. and J. Schneider (2017). "Optimal Alternative Dispute Resolution". *mimeo*.

— (2018). "Persuading to Participate: Coordinating on a Standard". *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.

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

Brown, S., C. Cervenak, and D. Fairman (1998). *Alternative Dispute Resolution Practitioners Guide*. Tech. rep. Conflict Management Group.

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.

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.

Correia-da-Silva, J. (2017). "Trembling mechanisms". *mimeo.*

Crémer, J. and R. P. McLean (1988). "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions". *Econometrica* 56, pp. 1247–1257.

Dworczak, P. (2017). "Mechanism Design with Aftermarkets: Cutoff Mechanisms". *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.*

Genn, H. (1998). *The Central London County Court Pilot Mediation Scheme.* Tech. rep. Department for Constitutional Affairs.

Gerardi, D. and R. B. Myerson (2007). "Sequential equilibria in Bayesian games with communication". *Games and Economic Behavior* 60, pp. 104–134.

Grossman, S. J. and O. D. Hart (1983). "An analysis of the principal-agent problem". *Econometrica*, pp. 7–45.

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

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.

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

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

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

Prescott, J. J., K. E. Spier, and A. Yoon (2014). "Trial and Settlement: A Study of High-Low Agreements". *Journal of Law and Economics* 57, pp. 699–821.

Siegel, R. (2014). "Asymmetric all-pay auctions with interdependent valuations". *Journal of Economic Theory* 153, pp. 684–702.

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

— (2007). "Litigation". *Handbook of law and economics* 1, pp. 259–342.

Strulovici, B. and R. Siegel (2018). "Judicial Mechanisms". *mimeo.*

Sugaya, T. and A. Wolitzky (2017). "Revelation Principles in Multistage Games". *mimeo.*

Vasserman, S. and M. Yildiz (2018). "Pretrial Negotiations Under Optimism". *RAND Journal of Economics* forthcoming.

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

Zheng, C. (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.*