Managing A Conflict*

Benjamin Balzer† Johannes Schneider‡

November 3, 2017

Abstract

We characterize the design of mechanisms aiming to settle conflicts that otherwise escalate to a costly game. Participation is voluntary. Players have private information about their strength in the escalation game. The designer fully controls settlement negotiations but has no control over the escalation game. We transform the mechanism design problem of conflict management to the information design problem of belief management conditional on escalation. The transformed problem identifies how the properties of the escalation game influence the optimal mechanism. Applying our results to two types of escalation games, we obtain qualitative differences driven by the game's sensitivity to information.

1 Introduction

Conflict management that aims to settle a conflict at little or no cost often operates in the shadow of some default resolution mechanism. Should an intermediary fail to settle the conflict, it escalates to a fight, that is, a non-cooperative game beyond the intermediary's control. If the conflicting parties hold private information regarding their ability in that escalation game, continuation strategies depend on the information obtained during conflict management. Thus, when designing conflict management, the intermediary has to take into account that information revelation during conflict management influences players' behavior in the escalation game.

Consider, for example, alternative dispute resolution (ADR) in legal disputes aiming to find out-of-court settlement solutions. Should ADR fail to settle the conflict, the conflict escalates, and disputants revert to formal litigation in court. The procedural rules of litigation are beyond control of the designer of the ADR mechanism. However, the designer's choices influence how informative the process of ADR is to disputants.

*We thank Eddie Dekel, Philipp Denter, Martin Dumav, Daniel Garrett, Johannes Hoerner, Antoine Loeper, Benny Moldovanu, Volker Nocke, Martin Peitz, Thomas Troeger, Christoph Wolf, Asher Wolinsky, and audiences in Bonn, Mannheim, Madrid, Munich, Warwick, Melbourne, Geneva, Palma, Lisbon, Faro, and Maastricht 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, jschneid@eco.uc3m.es

Naturally, strategic choices in litigation following a failed ADR attempt depend on the information obtained during ADR. Thus, the designer must account for the effect of information revelation *during* ADR on disputants' behavior *after* escalation. A disputant who, in turn, decides whether to participate in an ADR mechanism takes into account how information revelation affects both her and her opponent's continuation strategy.

Examples of conflict management in the shadow of an escalation game abound. Besides legal disputes they include a mediator in collective bargaining that operates in the shadow of a strike, peace negotiators concerned about escalation to war, or negotiations over trade agreements limited by a sovereign country's right to impose tariffs.

In this paper, we characterize optimal conflict management when two players dispute over the allocation of a given pie. Conflict management's aim is to prevent escalation to a game beyond the designer's control. Continuation strategies – and hence expected outcomes – depend on beliefs about the distribution of ability in the escalation game. We address the following questions for a broad class of escalation games: *Which aspects of the escalation game are most important for optimal conflict management? What is the role of information in the relation between conflict management and the escalation game? What influences players' strategic choices before, during and after conflict management?*

We characterize optimal conflict management for a broad class of escalation games. The designer has to consider how her mechanism influences players' strategic choices should the conflict escalate. Both the structure of the escalation game and information revelation during conflict management determine the players' choices.

To gain intuition for the designer's problem, consider the ADR example from above. In formal litigation disputants provide evidence to convince a judge or a jury. Evidence provision is costly and destroys part of the joint surplus. A disputant's cost function is her private information. Absent ADR, a low-cost disputant expects to succeed in litigation easily. She requires a *favorable settlement* to participate in ADR. Absent escalation, a high-cost disputant may bluff and claim low cost since no evidence is provided and actual cost are irrelevant. Occasional *escalation* to litigation *deters bluffing.*

To understand how continuation strategies react to the information structure, consider ADR that cannot guarantee full settlement, that is, ADR sometimes leads to escalation. The structure of ADR and the event of escalation jointly reveal information about the distribution of cost functions. Disputants use this information to adjust their continuation strategies. Suppose a plaintiff *expects* that on average the defendant has a much lower cost than she does. Then, her chances of winning are low and she reduces evidence provision in litigation to save cost.

Thus, part of the mechanism design problem is to design the information structure at the beginning of the escalation game. That information design problem interacts with the mechanism in a non-trivial way. A player has an incentive to use conflict management only to *plant false information* into her opponent's mind. If, for example, a high-cost plaintiff convinces the defendant that she has low cost, the defendant may respond by reducing the amount of evidence. The plaintiff then exploits this change in behavior. Naturally, optimal conflict management takes these incentives into account.

Results and Implications. Our main result, Theorem 2, establishes a duality between the primal mechanism-design problem of conflict management and a dual information-design problem of belief management. It characterizes optimal conflict management as the solution to optimal belief management in the escalation game.

In the dual problem, the designer chooses the information structure in the event of escalation. Thus, she picks the "prior" of the escalation game. Her objective is to maximize the sum of two measures, both directly defined on the properties of the escalation game. The first measure relates to effectiveness of screening. It quantifies the discrimination between types, that is, the importance of a player's exogenous type on continuation payoffs as a function of the information structure. The second measure determines how attractive it is to participate in conflict management. It is aggregate welfare conditional on escalation.

The characterization has several advantages. First, it significantly simplifies the problem. The primal mechanism-design problem involves non-trivial effects of the design choices on behavior in the escalation game. Design choices influence the settlement stage, the escalation stage, and the probability of either occurring. The dual information-design problem instead concentrates on the event of escalation. The information structure solving the dual problem is a sufficient statistic for the (reduced-form) mechanism. That is, there is a one-to-one mapping between possible solutions to the information design problem and candidate mechanisms. As a byproduct, the formulation disentangles the information-revelation part from the "game-design" part of the problem. The separation identifies the channel that connects the escalation game with the mechanism.

Second, the characterization highlights the economic channel through which optimal conflict management operates. The role of the escalation game is twofold. It serves as a last resort to the designer to verify claims during conflict management by making private information payoff relevant. In addition, it ensures that players expect to obtain some payoff even if conflict management fails to settle the conflict. The dual objective identifies, and quantifies the importance of these two effects as a function of the escalation game's properties. The effect of the information structure on the discrimination measure captures the first effect, that on the welfare measure captures the second.

Finally, the characterization illustrates that understanding the role of information in the escalation game is crucial for the design of optimal conflict management.

We demonstrate the last point by comparing two different escalation games, simple lotteries and contests. Both are common in the literature on *conflicts*, yet the literature on *conflict management* almost exclusively focuses on simple lotteries for tractability reasons. Applying Theorem 2 the solution is straightforward in both types of games. Yet, optimal conflict management differs drastically between the two.

A property of simple lotteries is that escalation destroys a fixed proportion of surplus.

In addition, the implicit assumption in such a game is that any type's *action choice* is constant in the information structure. That is, the distribution of the remaining surplus depends only on the given *type pair* and is thus linear in the information structure. Aggregate welfare is constant. We confirm the findings of the literature that the mechanism sorts type pairs into categories, promising weak pairings settlement while strong pairings often escalate. In addition, we provide an information-based explanation of this result. From the designer's point of view each type pair has some "virtual rent". The designer's problem reduces to assigning as much probability weight as possible to those pairs with high virtual rents providing the familiar sorting result.

The second type of game, a contest, is common to describe conflicts, yet largely ignored when it comes to conflict management. Not only is our information-based approach tractable, but it also demonstrates that results are in stark contrast to the lottery case. The main reason is that equilibrium action choices in contests are sensitive to the information structure. We show that governing the relation between information revelation and continuation strategies is of first-order importance to the designer.

The designer ensures that players cannot influence their information set through behavior within conflict management. She promises all types an information set that is independent of their own behavior. As a byproduct, the mechanism cannot promise settlement with probability one for any type profile.

Moreover, while (symmetric) lotteries lead to symmetric solutions, symmetry in beliefs is never optimal in a contest. Instead there is always one player that appears stronger *conditional on escalation.* The reason for this *endogenous asymmetry* is that aggregate welfare is larger if contestants enter the contest asymmetrically.

Structure of the Argument. We proceed in steps towards our result. We describe the setup of the model in Section 2. In Section 3 we first derive a condition, Proposition 1, that determines in which cases conflict management can guarantee a settlement solution. We show that the question whether full settlement is implementable reduces to: *Can the mechanism implement an equal split of surplus?*

Next, we consider cases in which full settlement is not implementable. Using binding constraints and properties of Bayes' rule, Theorem 1 establishes a one-to-one mapping between the information structure in case of escalation and a set of reduced-form mechanisms. For a given information structure, the mapping identifies the optimal mechanism implementing this information structure. Thus, effectively, Theorem 1 reduces the problem to: *What is the optimal information structure conditional on escalation?*

We proceed by transforming the designer's objective function. In Proposition 2, we state necessary and sufficient conditions when a reduced-form mechanism is implementable. We then use first-order conditions to develop an equivalent formulation of the problem, the dual. The dual is the sum of the two economic measures discussed above, discrimination and welfare. Both depend on the primitives of the escalation game and take the information structure at the event of escalation as their argument. Using the dual objective we characterize the optimum as the answer to the question: *Which information structure maximizes discrimination and welfare in the escalation game?*

In Section 4 we apply our results to simple lotteries and contests. In Section 5 we extend the model in various dimensions and conclude in Section 6.

Related Literature. We build on the theory of bilateral trade mechanisms initiated by Myerson and Satterthwaite (1983). Similar to Compte and Jehiel (2009), we assume a division of a pie as an outcome, a budget constraint mechanism, and continuation payoffs after escalation that depend on the information structure. We include interdependent valuations in the sense of Jehiel and Moldovanu (2001) that lead to an information externality. Two aspects differ, however. First, we allow for a continuation payoff that is non-linear in beliefs.[1] We show that the departure from linearity has important consequences on optimal design. Second, our focus is different as we also characterize the second-best mechanism. To do so, we depart from standard solution approaches and consider players' information sets. Thereby, we characterize and interpret the optimal solution using the properties of the escalation game directly.

Close in spirit are Philippon and Skreta (2012) and Tirole (2012) who study the informational externality of a bailout mechanism on future market behavior. In line with our approach, they consider a model in which the design of the mechanism influences the interpretation of observed behavior and thus subsequent choices in the market. A similar approach is taken by the literature on aftermarkets (Atakan and Ekmekci, 2014; Dworczak, 2017; Lauermann and Virág, 2012; Zhang, 2014) that considers how informational externalities of a mechanism influence behavior in the aftermarket. Importantly, aftermarket competition is a game between the winner of the initial mechanism and a set of new players not participating in the initial mechanism. Therefore, the design of the mechanism affects the information structure only one-sidedly and beliefs are type-independent by construction. In our model, the same players meet in the mechanism and in the escalation game. Thus, the mechanism's design influences the information of all players and type-dependent beliefs are possible.

We contribute to the literature on conflict preemption. In line with Zheng (2017), we derive a necessary and sufficient condition for full settlement. Beyond that, our analysis addresses the optimal second-best mechanism in case the condition is not met. The existing literature on such second-best mechanisms (Bester and Wärneryd, 2006; Hörner, Morelli, and Squintani, 2015; Spier, 1994) ignores the influence of conflict management on continuation strategies, mainly for tractability.

We nest these existing models. However, we show that economic forces differ substantially if continuation strategies are information sensitive. Thus, our approach to conflict management is more general including cases in which information revelation af-

[1] Fieseler, Kittsteiner, and Moldovanu (2003) and the literature on second-best conflict preemption (Bester and Wärneryd, 2006; Hörner, Morelli, and Squintani, 2015; Spier, 1994) considers similar models. In all of those, continuation payoffs are linear in beliefs.

fects expected outcomes non-linearly through changes in players' continuation strategies. We identify, and quantify the mechanism's effect on subsequent behavior complementing Meirowitz et al. (2017) who study investment prior to conflict management.

Conflict management is inherently a mechanism design problem, but we show that a dual information-design problem exists. The information design problem is simpler to solve by straightforwardly applying methods from Bergemann and Morris (2016).

2 Model

Grand Game. There are two ex-ante identical, risk-neutral players competing for a pie worth 1 to each player. There are two ways to solve the dispute. The players can engage in an exogenously given game of conflict, $\mathcal{V}$, such as litigation, or they can seek conciliation via some form of conflict management, $\mathcal{CM}$, such as alternative dispute resolution (ADR). Conflict management takes place only if *both* parties agree to participate. Otherwise they non-cooperatively play $\mathcal{V}$. The veto structure captures, for example, the idea of the rule of law providing a constitutional right to enforce a lawsuit.
Conflict Management. Conflict management is a mechanism proposed by a non-strategic third party, the designer, at the beginning of the game. It leads either to a settlement solution, $\mathcal{Z}$, or to escalation, $\mathcal{G}$. A settlement solution awards player i a share of the pie, x_i, such that $x_1 + x_2 \leq 1$.[2] Escalation triggers a non-cooperative game $\mathcal{G}$. The two games $\mathcal{G}$ and $\mathcal{V}$ may be identical, but do not have to be.[3] In any case, $\mathcal{G}$, too, is beyond the designer's control. That is, once the conflict escalates the designer controls neither players' action choices nor the payoff rule that $\mathcal{G}$ implements. Conflict management thus describes a mechanism design problem within a greater strategic environment beyond the designer's control, e.g., designing ADR within the legal system.
Initial Information Structure. Each player i is endowed with type θ_i independently drawn from the same distribution over $\Theta = \{1, 2, ..., K\}$. The probability of being a specific type θ_i is $p(\theta_i) > 0$. The type is payoff relevant in $\mathcal{V}$ and $\mathcal{G}$, but not under settlement. That is, we assume that players have identical preferences for the good, but may differ in their *ability* to play the non-cooperative games. In a legal dispute θ_i represents the cost of providing formal evidence in litigation which is unrelated to a disputant's value of winning the case. All, but the realization of θ_i 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 $\mathcal{G}$. Let $A^{\mathcal{G}} \subset \mathbb{R}^2$ describe the space of joint actions in $\mathcal{G}$. We assume that the escalation game is more inefficient than settlement. That is, its payoff function is a mapping $(u_1, u_2) : \Theta^2 \times A^{\mathcal{G}} \to (-\infty, 1]^2$ with $u_1 + u_2 \leq 1$. Similarly, the payoff function for the veto game $\mathcal{V}$ is $(v_1, v_2) : \Theta^2 \times A^{\mathcal{V}} \to (-\infty, 1]^2$ with $v_1 + v_2 \leq 1$.

[2] If the pie is indivisible, we interpret x_i as probability of winning the whole pie.

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

We assume that u_i and v_i are weakly decreasing in θ_i capturing the interpretation of types as (marginal) *cost-of-effort parameters* in both $\mathcal{V}$ and $\mathcal{G}$. Throughout the paper we use the terms stronger (weaker) to indicate lower (higher) values of θ_i. For simplicity we assume symmetry, that is, $u_i(\theta_i, \theta_{-i}, a_i, a_{-i}) = u(\theta_i, \theta_{-i}, a_i, a_{-i})$ and $v_i = v$.

Timing. After learning their type, players simultaneously decide whether to ratify or veto conflict management. If at least one player vetoes, her identity is revealed, players play $\mathcal{V}$ and payoffs realize. If both players ratify conflict management, the mechanism is played and either implements a settlement solution, or the conflict escalates to $\mathcal{G}$. For simplicity we focus on arbitration and assume full commitment upon ratification. This assumption implies in particular that the mechanism can directly or indirectly trigger the escalation game by creating an environment sufficiently hostile so that players refuse to settle. We discuss ways to relax this assumption in Section 5.

Solution Concept. We use perfect Bayesian equilibrium (Fudenberg and Tirole, 1988). We aim to find the mechanism that maximizes the probability of settlement. The choice of the objective is driven by applications in which either escalation implies externalities on the society, e.g., by limiting access to the legal system, or quality of conflict management is judged by "cases solved" rather than by details of settlement contracts. In addition, our objective emphasizes that the designer's main concern is about behavior in the escalation game, despite her being genuinely agnostic about the resulting outcomes.[4]

Key Modeling Choices. A key feature of our model is that a player's strength under escalation is orthogonal to her preferences regarding outcomes. The motivation behind this modeling choice lies in the institutional specifics of the environment. Often, the institutional environment requires certain skills that are not directly connected to the value of winning. For example, in litigation, there are typically aspects determining access to evidence that are uncorrelated with the value of winning the case. All results extend to settings that include additional private information about preferences.

A second important assumption is that all settlement solutions are based on soft information only. In reality, such mechanisms are frequent as they are the least costly solutions. In case of ADR, the outcome function of any more complicated ADR mechanism that includes provision of evidence *within* ADR can be subsumed in the escalation stage with our mechanism focusing on initial exchange of soft information only.

Finally, the assumption that the mechanism 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 designer* is anticipated by parties in advance. In reality, however, out-of-court settlement solutions result mainly from interaction with retired judges or mediators who are trained in conflict management and who provide their services repeatedly having an incentive to commit.

[4] Maximizing settlement is *not equivalent* to maximizing utilitarian welfare. Our main insights remain under such an alternative objective. See Section 5 for a discussion.

3 Analysis

3.1 First-Best Benchmark: Full-Settlement Solutions

We first provide conditions for a full-settlement mechanism. Invoking the revelation principle we restrict attention to direct revelation mechanisms. Any such mechanism implements a decision rule based on players' type reports. The decision rule is incentive compatible and satisfies participation constraints. Types are payoff irrelevant under settlement. Thus, any mechanism completely avoiding both $\mathcal{V}$ and $\mathcal{G}$ is a pooling mechanism.

Any pooling mechanism implements a particular sharing rule (x_1, x_2) independent of the players' reports. Whether such a rule satisfies the participation constraints depends on the expected payoff from vetoing the mechanism. A player's optimal strategy in $\mathcal{V}$ maximizes her expected payoff over action choices conditional on her belief about the opponent's type *and* corresponding strategy. The entire information structure becomes relevant, since the opponent's strategy is a function of the opponent's beliefs.

Suppose player i vetoes the mechanism. Then, she learns nothing about her opponent and continues to use the common prior p to evaluate $-i$'s type distribution. The non-vetoing player $-i$, to the contrary, may learn from i's veto. She is going to use the probability function ρ^V over i's type. Via the "no-signaling-what-you-don't-know" condition of perfect Bayesian equilibrium ρ^V is independent of θ_{-i}.[5] The information structure after i vetoes is (p, ρ^V). Behavior in the continuation game after a veto forms a Bayes Nash equilibrium given (p, ρ^V). Player i's expected continuation payoff is

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

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 Zheng (2017). It determines a necessary and sufficient condition for a full-settlement mechanism.

Proposition 1 (Full Settlement Mechanisms)**.** *The optimal conflict management mechanism guarantees full settlement if and only if there is a probability mass function $\tilde{\rho}$ over the type space such that $V_i(1, (p, \tilde{\rho})) \leq 1/2$.*

The result combines three constraints. Settlement has no screening power, thus the designer offers a pooling mechanism; conflict management is budget constraint, thus settlement divides the pie; and participation is voluntary, thus the designer incentivizes players via a sufficiently large share. Players are ex-ante symmetric, and a first-best mechanism exists only for $\mathcal{V}$ sufficiently costly. Proposition 1 reduces existence of the first-best to the simple question: *Can we implement an equal split of the pie?*

[5] In the first-best benchmark this belief is off the equilibrium path. In principle nothing changes if a veto occurs on the equilibrium path, but that ρ^V is derived via Bayes' rule.

3.2 Second-Best Mechanisms: The Designer's Problem

We now turn to analyzing 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 designer's problem as a primitive. The designer has no direct control over the players' decisions, given (ρ_i, ρ_{-i}). As we are going to see, a good knowledge of the game's equilibrium properties is crucial when designing the mechanism. Hence, we make our 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 $\mathcal{B}$ and satisfies the following.
(HC) Upper hemicontinuity. V_i is upper hemicontinuous in (ρ, ρ').
(S) Symmetry. $V_1(\theta, (\rho, \rho')) = V_2(\theta, (\rho', \rho))$ for any θ, (ρ, ρ').
(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.

Property (HC) implies that the designer'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 1 does not apply. The property (CONV) (together with Assumption 2 made below) avoids tedious case distinctions by ensuring that no player vetoes the optimal mechanism. We extend our analysis relaxing (CONV) in Section 5.

If full settlement is not implementable and no player vetoes on the equilibrium path, second-best conflict management uses escalation as a screening device. Once the conflict escalates, the designer's influence ceases and $\mathcal{G}$ is played non-cooperatively. We apply the revelation principle of Myerson (1982) to account for the strategic interaction after escalation. Thus, it is without loss of generality to restrict the set of mechanisms to functions mapping type reports into (i) a probability that the conflict escalates, γ, (ii) a sharing rule under settlement, X, and (iii) an additional public signal, i.e., a random variable, Σ, with realization σ. That is, conflict management is a mapping

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

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. Hence, to maintain simplicity we suppress signals notationally in the exposition whenever convenient. All formal 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.[6]

Let θ_i's value from participating in the mechanism and reporting m_i be $\Pi_i(m_i, \theta_i)$. Further, let the probability that conflict management breaks down and the conflict

[6]Conceptually, 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.

escalates be $Pr(\mathcal{G})$. Then, the optimal mechanism with full participation solves

$$\begin{aligned} &\min_{(\gamma,X,\Sigma)} Pr(\mathcal{G}) \text{ s.t.} \\ &\Pi_i(\theta_i;\theta_i) \geq \Pi_i(m_i;\theta_i), \qquad \forall (m_i;\theta_i) \in \Theta^2,\ i \in \{1,2\} \\ &\Pi_i(\theta_i;\theta_i) \geq V_i(\theta_i,(p,\rho^V)), \qquad \forall \theta_i \in \Theta,\ i \in \{1,2\}. \end{aligned} \tag{P$_{min}$}$$

The first set of constraints ensures incentive compatibility, the second full participation.[7] The value from participation, Π_i, depends on behavior in conflict management assuming optimal behavior in the continuation game. Next, we add more structure to Π_i using the equilibrium properties of the escalation game.

3.3 Incentives and Information

There is a conceptual difference between the information sets in games $\mathcal{V}$ and $\mathcal{G}$. Different to the veto game, a player knows her own report at the beginning of $\mathcal{G}$, and thus enters with a private history of past action. Hence, (equilibrium) beliefs are not the same across types. We describe the information structure using a $K \times 2$ matrix of probability distributions. An element of this matrix, $\beta_i(\cdot|\theta_i)$, is θ_i's *individual belief.* The individual belief describes θ_i's information about player $-i$'s type at the start of the (on-path) escalation game. The collection of all types' beliefs is a K-dimensional vector, $\boldsymbol{\beta}_i$, of probability distributions. The on-path information structure, $\mathcal{B} := (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, contains all possible on-path beliefs. We refer to $\mathcal{B}$ as a *belief system.* By the revelation principle, the escalation rule γ determines $\mathcal{B}$, and $\mathcal{B}$ is common knowledge.

At the beginning of the grand game the designer is uninformed about the players' type. Thus, γ conditions only on *reports.* Consequently, individual beliefs differ with their report. A player's private information at the beginning of the escalation game consists both of her *true payoff type* and her *reported type.* She expects no more than

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

where $G(a_{-i}|\theta_{-i},\mathcal{B})$ is θ_{-i}'s distribution over equilibrium actions.[8] We impose simplifying assumptions on $\mathcal{G}$ for notational convenience.

Assumption 2. U_i exists 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, $U_i \geq V_i$.[9]

The last part of Assumption 2 implies that players are not *exogenously* punished for attending conflict management. Together with Assumption 1 it is sufficient for the

[7] Assumption 2 below ensures that the optimal mechanism indeed involves full participation.

[8] 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.

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

revelation principle to hold. That is, an optimum with full participation always exists.

The continuation payoff (U) illustrates the main features of our model. First, the individual belief *depends on the type report only.* Second, the belief about each type's action depends on the entire on-path belief system $\mathcal{B}$. On the equilibrium path each player expects her opponent to report truthfully and to follow an equilibrium strategy in the escalation game. Third, the designer influences the continuation game only through $\mathcal{B}$ which itself is determined by the escalation rule. All other elements in (U) are beyond the designer's control, in particular the functional form of U_i. Therefore we treat U_i as a primitive to the problem.

Finally, (U) highlights the main difference between veto and escalation. Players cannot always replicate escalation in the veto game even if $V_i = U_i$. Any information set post-veto only depends on the vetoing player's *identity.* The escalation rule γ can implement a richer set. In particular, it induces individual beliefs that depend on *identity and report.* Thus, different types may have different individual beliefs. For the special case that beliefs are constant in m_i, the continuation payoff U_i is constant in m_i, too.

To address the incentive problem, let us consider the expected payoff from participating in conflict management. For now, assume that the mechanism releases no information beyond the escalation decision, that is, no additional public signal is sent. Then, the expected payoff from participation is

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

where $\gamma_i(m_i) := \sum_{\theta_{-i}} p(\theta_{-i})\gamma(m_i,\theta_{-i})$ is the probability of escalation when reporting m_i. Conceptually, the expected payoff can be split in two parts. The settlement value, z_i, depending on a player's report only, and the escalation value, $y_i(m_i;\theta_i)$, which is a function of the report and the payoff type of the player.

To illustrate the novelty in our model, consider two adjacent types θ_i, $\theta_i{+}1$ both reporting θ_i during conflict management. The difference in their expected payoffs is

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

We refer to the difference $D_i(\theta_i;\theta_i,\mathcal{B})$ as player θ_i's *ability premium.* It is the difference between player i's on-path continuation payoff and that of the next strongest player mimicking θ_i. It measures how much of that payoff is due to θ_i's payoff type. Different to the existing literature, a change in $\mathcal{B}$ affects the ability premium non-linearly. Misreporting alters both the beliefs and as a consequence the deviator's action. Moreover, misreporting always provides an informational advantage to the deviator. Deviators remain undetected by definition and operate under superior knowledge. Their action is a best response to the opponent's on-path strategy. The reverse, however, does not need to hold as the opponent expects equilibrium play.

The expected payoff from conflict management, (1), also illustrates the main difficulty of the problem. Any choice of γ has a non-linear effect on U_i via $\mathcal{B}$. Standard methods are thus not directly applicable and we proceed with a change of variable.

3.4 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 the conflict management problem into a problem of managing beliefs in the escalation game. The transformation not only reduces complexity, but 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 directly via the properties of the escalation game $\mathcal{G}$. Belief management identifies the channel through which information in the escalation game enters the considerations of the designer.

Two steps yield that 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 designer can implement given the prior. Consistency allows for $K + 1$ independent belief distributions.[10]

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 any realization of the public signal Σ into a consistent belief system.

The random variable $\mathcal{B}(\Sigma)$ links the realization σ directly to a consistent belief system. Thus, if σ is induced with $Pr(\sigma)$ via Σ, so is the corresponding consistent belief system. Any (stochastic) mechanism $\mathcal{CM}$ trivially induces a random consistent belief system. Our first theorem shows 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.

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

Theorem 1. *Under Assumption 1 and 2 the set of feasible random belief systems $\mathcal{B}(\Sigma)$ is well-defined. 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 $CM : \mathcal{B}(\Sigma) \mapsto (z, \gamma)$ 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 with respect to the escalation rule. Escalation value, y_i, and escalation probability, γ_i, are homogeneous of degree 1 with respect to 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 $\mathcal{G}$ 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 $\overline{\alpha}\gamma(\theta_1, \theta_2) \leq 1$ for all (θ_1, θ_2). Then, the rule $\overline{g}_{\mathcal{B}} := \overline{\alpha}\gamma$ maximizes escalation and is thus the worst 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 compatible with an incentive compatible mechanism $\mathcal{CM}$ satisfying participation constraints.

Observation 3. For any type θ_i with positive settlement value it is without loss to assume that either an incentive constraint, or the participation constraint binds in the second-best mechanism. Moreover, given $\mathcal{B}$, 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 designer is budget constraint under settlement. The second part is immediate when combining a player's expected payoff from a mechanism, expression (1), with Observation 2. Observation 3 implies that $\mathcal{B}$ captures the entire non-linear part of the constraint. Furthermore, given $\mathcal{B}$ the set of constraints consists of $2K$ linear equations with $2K + 1$ unknowns, the $2K$ settlement values and the scalar α. To close the problem we thus need one more equation. We use the resource constraint of the designer at the level of expected settlement payments,

$$\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \leq 1 - Pr(\mathcal{G}). \tag{2}$$

Observation 4. A reduced-form mechanism is feasible only if it satisfies (2).

The resource constraint, (2), is an immediate consequence of the designers budget constraint. The designer 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 $1 - Pr(\mathcal{G})$. Since Proposition 1 does not apply, the resource constraint binds at the optimum and is our final equation. For any consistent $\mathcal{B}$ there is a unique tuple (z^*, α^*) satisfying all 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 $CM : \mathcal{B} \mapsto (z, \gamma)$ that identifies a *unique candidate* (z, γ) for any $\mathcal{B} \notin \{\mathcal{B}\}_\emptyset$ and points to the origin otherwise. The function CM is continuous in the interior of the support and given by

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

Discussion of Theorem 1. The conflict management problem (P_{min}) contains both a mechanism design and an information design part. The designer decides on the game form and acts as a mechanism designer. Yet if she invokes escalation, her actions are restricted to the distribution of information. Thus, she acts as an information designer in the escalation game. However, prior to distribution she has to elicit the information by designing the settlement game. 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 CM determines whether that belief system is affordable. Moreover, CM determines the escalation rule that minimizes the price (in terms of lost settlement) 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?*

3.5 A Dual Problem to Optimal Conflict Management

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 in turns.

First, we translate the results of Border (2007) to our setup and state necessary and sufficient conditions to implement a reduced-form mechanism (z, γ) via some $\mathcal{CM}$. Let

$Q \subset \Theta^2$ be any subset of the type space and define $Q_i := \{\theta_i | \exists \theta_{-i} : (\theta_i, \theta_{-i}) \in Q\}$ and $\tilde{Q} := \{(\theta_1, \theta_2) \in \Theta^2 | \theta_i \notin Q_i \text{ for } i = 1, 2\}$.

Proposition 2 (Border (2007))**.** *Take any reduced form mechanism (z, γ). An ex-post feasible X_i that implements z exists if and only if*

$$\sum_i \sum_{\theta_i \in Q_i} z_i(\theta_i) p(\theta_i) \leq 1 - Pr(\mathcal{G}) - \sum_{(\theta_1,\theta_2) \in \tilde{Q}} (1 - \gamma(\theta_1, \theta_2)) p(\theta_1) p(\theta_2), \quad \forall Q \subseteq \Theta^2. \quad \text{(GI)}$$

Proposition 2 imposes a set of additional constraints. Following Border (2007) we call this set the general implementation condition, (GI).

Next, we define a regular environment in our setting. We state sufficient conditions that guarantee that the following constraints bind at the optimum: (i) the strongest types' participation constraints, and (ii) local downward incentive constraints.

The first requirement is related to the veto game. Regularity implies an environment in which the strongest type is sufficiently privileged in the veto game, for example, because she occurs sufficiently rare. Then, a strong type expects to face a weak opponent. She is optimistic about her prospects outside conflict management. The mechanism has to compensate the strongest type, making her participation constraint binding. The second requirement is related to the escalation game. Roughly put, it requires that the weaker a player is, the more important is her fundamental type θ_i. Recall that the ability premium, $D_i(m; \theta_i, \mathcal{B}) = U_i(m; \theta_i, \mathcal{B}) - U_i(m; \theta_i{+}1, \mathcal{B})$, is the difference in the continuation payoff of two adjacent types θ_i and $\theta_i{+}1$ reporting the same type m.

Definition 4 (MDR)**.** The game $\mathcal{G}$ 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) $2 \sum_{\theta_i \in \hat{Q}} p(\theta_i) V_i(\theta_i, (p, \underline{\rho}^V)) < \sum_{\theta_i \in \hat{Q}} p(\theta_i)$, for any $\hat{Q} \subseteq \Theta$ and $\hat{Q} \neq \{1\}$, and

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

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

Regularity, in particular (MDR), imposes strong restrictions on the nature of the escalation game, but allows us to focus on the economics when setting up the dual problem. Main insights carry over to a more general environment. In Section 5 we discuss the irregular case and its relation to the characterization provided in this section.

Lemma 1. *Suppose Assumption 1 to 3 hold. At the optimum local downward incentive constraints bind for any type $\theta_i > 1$. In addition, for each player the only binding participation constraint is that of type 1.*

Combining Theorem 1 with Lemma 1 leaves the designer with a two-fold objective. First, she wants to provide a belief system that maximizes the aggregate surplus in the continuation game. The larger the aggregate surplus in the continuation game the larger the escalation values. The expected payoff from participation is the sum of both the escalation value and the settlement value. Therefore, a larger settlement value relaxes the participation constraint. Second, the designer provides a belief system that allows her to *discriminate* between types. The ability premium, D_i, measures the discriminatory power of a belief system at the beginning of $\mathcal{G}$. Thus, the designer has an incentive to choose a belief system that leads to a relatively high ability premium. A high ability premium reduces the information rent the designer has to pay to weaker types. A smaller information rent, in turn, reduces the need to screen via escalation.

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 $w(\theta_i){:=}\Big(1 - \sum\limits_{k=1}^{\theta_i} p(k)\Big)/\Big(p(\theta_i)\Big)$.

We now derive a dual problem to (P_{min}) for the case that additional signals are superfluous. The set C contains all constraints in problem (P_{min}). The set of unambiguously binding constraints, $C_R \subset C$, are those defined in Lemma 1. These constraints are incorporated in the objective of the dual. We define $C_F := (C \setminus C_R) \cup (CM(\mathcal{B}) \neq 0)$, where the latter restricts the set of belief systems to those implementable by a feasible reduced-form mechanism. These constraints are outside the dual's objective. The conditional distribution of a player's type after escalation, ρ, is the solution to the following system of linear equations: $\rho_i(\theta_i) = \sum_{\theta_{-i}} \beta_{-i}(\theta_i|\theta_{-i})\rho_{-i}(\theta_i)\ \forall \theta_i$.

The dual objective is the simple sum of an ex-ante expected measures of discrimination, $\mathbb{E}[\Psi_i|\mathcal{B}] = \sum\limits_{\theta_i} \rho_i(\theta_i)\Psi_i(\theta_i,\mathcal{B})$, and one of welfare, $\mathbb{E}[U_i|\mathcal{B}] = \sum\limits_{\theta_i} \rho_i(\theta_i)U_i(\theta_i;\theta_i,\mathcal{B})$,

$$\max_{\mathcal{B} \in \{\mathcal{B}\}_a} \sum_i \mathbb{E}[\Psi_i|\mathcal{B}] + \mathbb{E}[U_i|\mathcal{B}] \quad s.t.\ C_F. \tag{$\mathrm{P}^{\mathcal{B}}_{max}$}$$

Proposition 3 (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,* $(z,\gamma) = CM(\mathcal{B}^*)$*, and* $\mathcal{B}^*$ *solves* ($\mathrm{P}^{\mathcal{B}}_{max}$)*.*

Proposition 3 follows from rearranging the first-order conditions and Theorem 1. We summarize the argument assuming $\gamma(1,1) \geq \gamma(k,n)$, that is, the two strongest types are the most likely to escalate. The escalation rule at the optimum is $\gamma(1,1)\overline{g}_{\mathcal{B}}$, where $\overline{g}_{\mathcal{B}}$ is the worst escalation rule for $\mathcal{B}$. Using Lemma 1 we rewrite the settlement values

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

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

$$\sum_i \sum_{\theta_i} p(\theta_i)z_i(\theta_i) = 2V_i(1,(\rho,\underline{\rho}^V)) + \gamma(1,1)\mathcal{Q}(\mathcal{B}) \overset{!}{=} 1 - Pr(\mathcal{G}) = 1 - \gamma(1,1)R(\mathcal{B}) \tag{3}$$

which implicitly determines $\mathcal{Q}(\mathcal{B})$ and $R(\mathcal{B})$. Solving equation (3) for $\gamma(1,1)$ and plugging into the objective yields

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

Any $\mathcal{B}$ that solves (P_{min}), solves sup $\mathcal{Q}(\mathcal{B})/R(\mathcal{B})$. Multiply the latter by $\gamma(1,1)/\gamma(1,1)$ and substitute back for $\gamma(1,1)\mathcal{Q}(\mathcal{B})$. Bayes' rule implies $\gamma(1,1)R(\mathcal{B}) = Pr(\mathcal{G})$ and $\gamma_i(\theta_i)/Pr(\mathcal{G}) = \rho_i(\theta_i)/p(\theta_i)$. Finally, substituting for z_i and y_i yields ($\mathrm{P}^{\mathcal{B}}_{max}$).

We postpone further discussion of Proposition 3 and first state the general case including public signals. Concavification results from the Bayesian persuasion literature do not apply directly as they build on the assumption of type-independent beliefs. The belief system $\mathcal{B}$, to the contrary, specifies type-dependent beliefs. Instead we use a two-step procedure.[11] We define the mean of $\mathcal{B}(\Sigma)$ as $\overline{\mathcal{B}} := \sum_{\sigma\in\Sigma} Pr(\sigma)\mathcal{B}(\sigma)$.

Definition 7 (Admissible Means)**.** The mean, $\overline{\mathcal{B}}$, is in the set of admissible means, $\{\overline{\mathcal{B}}\}^a$ if $\mathcal{B}(\Sigma)$ satisfies the constraints C_F.

For any mean, the set of signal structures, Σ, such that $\mathcal{B}(\Sigma)$ satisfies C_F is $\mathcal{S}(\overline{\mathcal{B}})$. Given objective $\mathcal{O}$ the value of $\overline{\mathcal{B}} \in \{\overline{\mathcal{B}}\}^a$ is

$$\mathcal{W}(\overline{\mathcal{B}}, \mathcal{O}(\mathcal{B}(\sigma))) := \max_{\Sigma\in\mathcal{S}(\overline{\mathcal{B}})} \sum_{\sigma\in\Sigma} Pr(\sigma)\mathcal{O}(\mathcal{B}(\sigma)). \tag{4}$$

Using (4) we generalize ($\mathrm{P}^{\mathcal{B}}_{max}$) to,

$$\max_{\overline{\mathcal{B}}\in\{\overline{\mathcal{B}}\}^a} \mathcal{W}\left(\overline{\mathcal{B}}, \sum_i \Big(\mathbb{E}[\Psi_i|\mathcal{B}(\sigma)] + \mathbb{E}[U_i|\mathcal{B}(\sigma)]\Big)\right). \tag{$\mathrm{P}^{\mathcal{B}(\Sigma)}_{max}$}$$

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

Every consistent belief system specifies a value of the objective. The value of the objective is no sufficient statistic for the optimal belief system because of additional binding constraints. Furthermore, we cannot solve the problem by choosing a signal structure such that $\{\mathcal{B}(\sigma)\}_\sigma$ spans the concave closure of the objective. The reason such a concavification approach fails is that $\{\mathcal{B}(\sigma)\}_\sigma$ spanning the concave closure may violate the binding constraints. Instead we introduce the lottery-means approach, which has no economic meaning per se. Mathematically, however, it exploits the linearity of equation (4) representing the maximization problem in a concise manner.

As a corollary to Theorem 2 we state a sufficient condition to when public signals are superfluous and Proposition 3 applies directly.

Corollary 1. *Consider the solution to* ($\mathrm{P}^{\mathcal{B}}_{max}$) *ignoring* C_F. *If this solution does not violate* C_F, *the optimal signal structure is degenerate.*

[11] In the supplementary material, appendix E, we use a Lagrangian approach and show that the optimal solution can be found by maximizing a concave closure of the supporting Lagrangian function.

The intuition for Corollary 1 is straightforward. Via Theorem 1 the most relevant constraints are included in the objective of ($\mathrm{P}^{\mathcal{B}}_{max}$). For a given information structure, signals might improve by linearizing convexities at the optimum (Aumann and Maschler, 1995). Convexities do not exist if the solution to ($\mathrm{P}^{\mathcal{B}}_{max}$) satisfies all remaining constraints by definition of a maximum. Thus, signals are superfluous. That is, instead of implementing a spread over information structures, the designer can implement the most preferred information structure with probability 1.

Discussion of Theorem 2. Theorem 2 provides a tractable dual to the conflict management problem. The economic interpretation of its objective is intuitive. Mimicking behavior has two effects on the continuation game: The deviator (i) inherits the posterior distribution over the opponent's types from the mimicked type, and (ii) gains an informational advantage being the only one aware of entering an off-path game. The deviator is not forced to adopt the strategy of the mimicked type in the continuation game, and can freely adjust her strategy. Choices remain unresponded by the opponent who plays *as if* she were on the equilibrium path. The deviator's information advantage reduces discrimination after escalation providing an additional incentive to deviate.

The problem ($\mathrm{P}^{\mathcal{B}}_{max}$) has a direct analogue to that of an optimal auction (Myerson, 1981). The main difference is that – although types are ordered due to Assumption 2 and 3 – the term $\Psi(\theta_i, \mathcal{B}) + U_i(\theta_i; \theta_i, \mathcal{B})$ is non-linear in the designer's choice. Consequently and different to the auction design problem, it is not possible to derive a simple characterization of the optimal mechanism without considering the details of $\mathcal{G}$.

Economically the dual problem provides important implications. First, to save on resources the designer reduces inefficiencies in the escalation game despite being agnostic about outcomes after escalation.

Second, combining Theorem 1 and 2 via ($\mathrm{P}^{\mathcal{B}(\Sigma)}_{max}$) we identify how details of the escalation game influence the optimal mechanism. More specifically, we define an information design problem. Its intuitive objective identifies and quantifies the designer's motives.

Tractability is guaranteed even in the presence of public signals. Theorem 2 describes a "backward-induction" approach to that general problem. First, we relax incentive constraints via information design in the continuation game assuming an omniscient designer. Second, we solve the core mechanism design problem that determines the price of information to the designer.

4 Application to Classical Games of Conflict

In this section we contrast characteristics of optimal conflict management in two classical models of (legal) disputes, a simple lottery and an all-pay auction. Variants of the former occur frequently in the existing literature on out-of-court settlements. We confirm the findings of the literature and provide an information-based explanation for them. The all-pay auction, although prominent in the literature on *conflict description*, is largely ignored in the context of *conflict management* for reasons of tractability. We show that the solution is straightforward and differs drastically from that of the simple lottery.

The comparison supports the relevance of our general result in two ways. First, the dual approach developed in Section 3 overcomes the tractability issue by directly operating *on the escalation game*. Second, using the dual problem ($\mathrm{P}^{\mathcal{B}}_{max}$) it becomes immediate that the two models have fundamental, qualitative differences.

In a simple lottery, optimal conflict management leads to a *type-dependent and symmetric belief* system, *without full support.* In particular, the two weakest types never face each other in the escalation game. Instead, optimal conflict management in an all-pay auction leads to a *type-independent, asymmetric* belief system *with full support.* In particular, all types face each other in the escalation game with positive probability.
Setting. Two players have a legal dispute. From an ex-ante point of view the case is "tight" and both players have the same probability of winning. With abuse of notation we assume a players' type realizes at 1 or $K > 2$. The ex-ante probability of being type 1 is p. Types represent merits of the case. Type 1 describes a "strong case" and type K describes a "weak case". We simplify the setup assuming that $v \equiv u$ such that $\mathcal{V}$ and $\mathcal{G}$ are identical given an information structure, and choose p low enough for Assumption 1 to hold. We refer to the realized state (θ_1, θ_2) as a *match* of types θ_1 and θ_2.

4.1 Simple Lotteries

Consider litigation as the following (reduced-form) game. Participation requires a fixed cost $0 \leq c < 1/2$ and a player's strategy choice depends on her own type only. Then, the continuation strategy is constant in the information structure and thus irrelevant.[12] To emphasize this effect, we suppress action choices in the payoff function. It is

$$u(\theta_i, \theta_{-i}) = \begin{cases} 1/2 - c, & \text{if } \theta_i = \theta_{-i} \\ 1/2 - c + \xi, & \text{if } \theta_i > \theta_{-i} \\ 1/2 - c - \xi, & \text{if } \theta_i < \theta_{-i}, \end{cases}$$

where $\xi > 0$ is the marginal payoff of a better case. Let $b_i(m_i) := \beta_i(1|m_i)$ be the probability mass on type 1, and $\mathbb{1}_1 = 1$ if θ_i=1 and 0 otherwise. Expected payoffs are

$$U_i(m_i; \theta_i, \mathcal{B}) = b_i(m_i)u(\theta_i, 1) + (1 - b_i(m_i))u(\theta_i, K) = 1/2 - c - b_i(m_i)\xi + \mathbb{1}_1\xi.$$

Optimal conflict management maximizes $\sum_i (\mathbb{E}[U_i|\mathcal{B}] + \mathbb{E}[\Psi_i|\mathcal{B}])$ subject to constraints C_F. K-types receive no virtual rent by definition, thus $\Psi(1, \mathcal{B}) = w(1)D_i(1; K, \mathcal{B})$.

Moreover, the ability premium, $D_i(1; K, \mathcal{B})$=$U_i(1; 1, \mathcal{B}) - U_i(1; K, \mathcal{B})$=$\xi$, is constant.

[12] In fact, we use the term *lottery* to emphasize that we could model the situation without the need of any formulation of a *game*. However, the application suggests existence of a *continuation game* in which each type's equilibrium strategy is invariant to the information structure. The two are observationally equivalent. A recent example of using this modeling technique is Hörner, Morelli, and Squintani (2015).

Hence, $\mathbb{E}[\Psi_i|\mathcal{B}] = \rho_i(1)w(1)\xi$ (linearly) increases in the post-escalation likelihood, $\rho_i(1)$. By design, the joint expected payoffs, $\sum_i \mathbb{E}[U_i|\mathcal{B}]{=}1{-}2c$, are independent of $\mathcal{B}$. Thus, the belief system that maximizes $\sum_i (\mathbb{E}[U_i|\mathcal{B}] + \mathbb{E}[\Psi_i|\mathcal{B}])$ is $(b_i(1) = 1, b_i(K) = 0)$ implying $\rho_i(1) = 1$. We denote this belief system by $\widetilde{\mathcal{B}}$.

The belief system $\widetilde{\mathcal{B}}$ *sorts* type profiles into the set of *easy-to-settle matches*, $\{(K, K), (1, K), (K, 1)\}$, and that of *difficult-to-settle matches*, $\{(1, 1)\}$. Under $\widetilde{\mathcal{B}}$ no type-1 player wants to mimic a type-K player. However, it is possible that the reduced-form mechanism implementing $\widetilde{\mathcal{B}}$ is not feasible. This happens if the worst escalation rule leading to $\widetilde{\mathcal{B}}$ does not satisfy the designer's resource constraint. By occasional escalation of $(1, K)$ and $(K, 1)$ the designer increases the escalation utility of type-1 players and, in turn, decreases their promised shares until the solution is resource feasible; (K, K)-matches remain settled for sure. The entire program is linear, and signals never improve.

Proposition 4. *Suppose the escalation game is the simple lottery described above. Optimal conflict management induces an on-path belief system in the escalation game with beliefs that are*

- *type-dependent, $b_i(1) \neq b_i(K)$, and*
- *degenerate for K-types, $b_i(K) = 1$.*

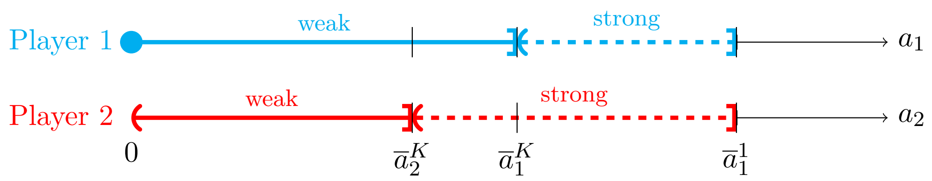
Furthermore, it is without loss of generality to assume that beliefs are symmetric*, $b_1(\theta_i) = b_2(\theta_i)$, and to ignore public signals.*

4.2 Contests

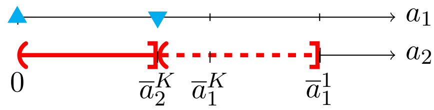
We contrast the results of simple lotteries with a model of litigation as a legal contest (cf. Posner, 1973, among others). The contest model assumes that disputants strategically exert costly effort to prove claims. The more effort a party exerts the better her legal argument and thus her chances to win the case. For the sake of the argument we consider the simplest model of a legal contest, the all-pay auction. We assume that each player decides how much effort, $a_i \geq 0$, to exert. We assume constant marginal cost of effort equal to type, θ_i. The player exerting most effort wins litigation. Payoffs are

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

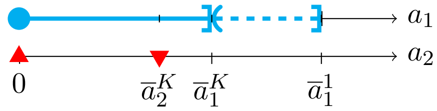
Contrary to simple lotteries, a player's payoff does not depend on the opponent's type directly, but on her *action choice*. We use $b_i(m_i) := \beta_i(1|m_i)$ and assume without loss of generality that $b_1(1) \geq b_2(1)$. Figure 1(a) graphs the unique monotonic mixed strategy equilibrium (see also Siegel, 2014). Types mix piecewise uniformly on disjoint intervals and at most one player has a mass point at 0. We use $F_i^{\theta_i}$ for θ_i's distribution over actions and $f_i^{\theta_i}$ for the density. Player i's winning probability at effort level a is



(a) Equilibrium effort support in the on-path continuation game. *All players and types play a mixed strategy. Solid lines denote the intervals of weak types, i.e., $\theta_i{=}K$, and dashed lines those of strong types, i.e., $\theta_i{=}1$. The equilibrium effort distribution is non-atomistic, apart from the mass point at 0 for player 1, type K.*



(b) Continuation effort after K_1 deviates. *Player 2 follows her equilibrium strategy. Player 1 chooses action ▲ if $b_2(1) > b_2(K)$ and ▼ if $b_2(1) < b_2(K)$.*



(c) Continuation effort after K_2 deviates. *Player 1 follows her equilibrium strategy. Player 2 chooses action ▲ if $b_1(1) > b_1(K)$ and ▼ if $b_1(1) < b_1(K)$.*

Figure 1: Continuation strategies for different histories if $b_1(1) \geq b_2(1) \neq b_2(K)$.

a function of her report and *her opponent's equilibrium strategy*. We denote it as the conditional distribution

$$F_{-i}(a_i|m_i) = b_i(m_i)F_{-i}^1(a_i) + (1 - b_i(m_i))F_{-i}^K(a_i).$$

Effort a_i after report m_i yields the continuation payoff

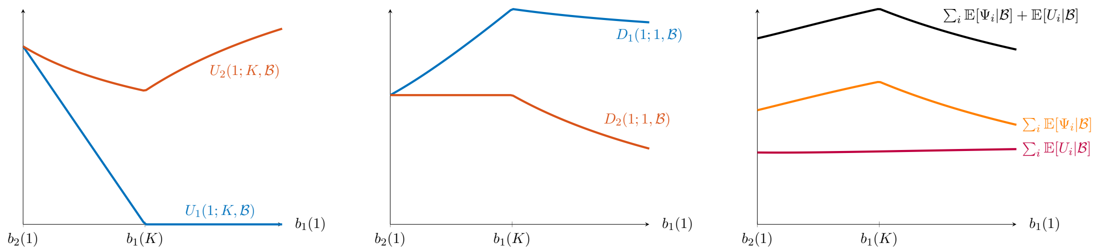
$$U_i(m_i; \theta_i, \mathcal{B}) = F_{-i}(a_i|m_i) - \theta_i a_i.$$

The common upper bound on optimal effort yields identical continuation payoffs for strong types. Applying the properties of best responses we obtain continuation payoffs

$$\begin{aligned}
U_i(1; 1, \mathcal{B}) &= F_1(\overline{a}_2^K|1) - \overline{a}_2^K \\
U_1(K; K, \mathcal{B}) &= 0 \\
U_2(K; K, \mathcal{B}) &= F_1(0|K) \\
U_1(1; K, \mathcal{B}) &= 0 + \mathbb{1}_{\leq}\left(F_2(\overline{a}_2^K|1) - K\overline{a}_2^K\right) \\
U_2(1; K, \mathcal{B}) &= F_1(0|1) + \mathbb{1}_{\leq}(F_1(\overline{a}_2^K|1) - K\overline{a}_2^K - F_1(0|1)),
\end{aligned} \tag{5}$$

where $\mathbb{1}_{\leq}$ is an indicator function with value 1 if $b_1(1) \leq b_1(K)$ and 0 otherwise.

To calculate the density functions $f_i^{\theta_i}$ we use the first-order condition of those types of player $-i$, whose equilibrium effort support intersects that of θ_i. On the equilibrium path player θ_{-i} is indifferent between a and any other action in her support. Hence, $f_i^{\theta_i}(a) = \theta_{-i}/(1 - b_i(\theta_i))$. The length of the different intervals in figure 1(a) is derived using all densities and the monotonic equilibrium structure. Finally, $F_1^K(0)$ is the residual mass for player 1, type K. We provide a full description of all relevant terms as functions of



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

Figure 2: Deviation in relation to $b_1(1)$. *Beliefs* $b_2(1), b_1(K)$ *are fixed,* $b_2(K)$ *adjusts endogenously ensuring consistency of* $\mathcal{B}$.

$\mathcal{B}$ in the appendix. Here, we focus on the conceptual intuition.

Consistency implies $\operatorname{sgn}(b_1(1)-b_1(K)) = \operatorname{sgn}(b_2(1)-b_2(K))$. Moreover, if $b_1(1) = b_1(K) \Leftrightarrow F_{-i}(\cdot|1) = F_{-i}(\cdot|K)$, that is, a deviating player does not gain superior information. A deviating type K faces the same information structure as a non-deviating player and is thus indifferent between any effort choice on the interval $(0, \overline{a}_2^K]$. As a result, the terms after the indicator $\mathbb{1}_{\leq}$ collapse to 0, and $U_i(1;K,\mathcal{B}) = U_i(K;K,\mathcal{B})$.

If instead $b_1(1) \neq b_1(K)$, a deviator still anticipates the opponent's type distribution correctly and best-responds using $F_{-i}(\cdot|m_i)$. The non-deviating player, to the contrary, is unaware of the deviation and plays *as if* on-path, using $F_i(\cdot|\theta_{-i})$. Her action is thus a best-response to faulty beliefs. We discuss the changes to the outcome case by case.

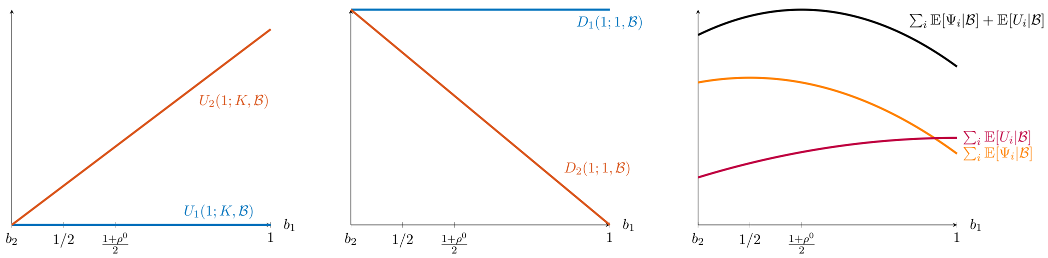
If $b_1(1) < b_1(K)$, a deviating type K expects a weaker opponent than on the equilibrium path. She reacts by putting full mass on $\overline{a}_2^K$. If $b_1(1) > b_1(K)$, she expects a tougher opponent. She reacts by putting full mass (close) to 0.[13] In the first case, she shifts mass upwards to obtain a higher probability of winning. In the second case, she shifts downwards to cut losses.

Deviation strategies are marked with the triangles in figures 1(b) and 1(c). Figure 2(a) graphs the corresponding deviation payoffs as a function of $b_1(1)$. Players' deviation payoffs are minimized if $b_1(1) = b_1(K)$, the case that eliminates their information advantage.

Obviously, the information advantage influences the ability premium. Unlike the lottery case, the ability premium is non-constant in $\mathcal{B}$,

$$
\begin{aligned}
D_1(1;1,\mathcal{B}) &= \begin{cases} F_1(\overline{a}_2^K|1) - \overline{a}_2^K & \text{if } b_1(1) > b_1(K) \\ F_1(\overline{a}_2^K|1) - F_2(\overline{a}_2^K|1) + (K-1)\overline{a}_2^K & \text{if } b_1(1) \leq b_1(K) \end{cases} \\
D_2(1;1,\mathcal{B}) &= \begin{cases} F_1(\overline{a}_2^K|1) - F_1(0|1) - \overline{a}_2^K & \text{if } b_1(1) > b_1(K) \\ (K-1)\overline{a}_2^K & \text{if } b_1(1) \leq b_1(K). \end{cases}
\end{aligned}
\tag{6}
$$

[13]Off-path continuation utilities are defined via the supremum. Although given our tie-breaking rule a deviating player 2 may have no best response, the supremum is reached as effort goes to 0.



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

Figure 3: Effect of asymmetry. *Increasing b_1 from $b_1 = b_2$ while holding beliefs type independent.*

Figure 2(b) depicts the ability premium as a function of $b_1(1)$. If a deviator expects to face a weaker opponent, i.e., $b_1(1) < b_1(K)$, an increase in $b_1(1)$ has a (weakly) positive effect on the ability premium. It reduces the size of the deviator's information advantage. If, however, a deviator expects a stronger opponent, the opposite is true. If $b_1(1)$ increases, then the deviator's information advantage increases. Moreover, if $b_1(1)$ increases, strong types expect tougher competition also on the equilibrium path. Thus, the ability premium shrinks.

The effect prevails in the expected virtual rent, $\sum_i \mathbb{E}[\Psi_i|\mathcal{B}]$, graphed in figure 2(c). The expected continuation welfare, $\sum_i \mathbb{E}(U_i|\mathcal{B})$, changes only marginally in $b_1(1)$. Hence, the optimal mechanism eliminates the information advantage for deviators. All players receive the same information regardless of their behavior within conflict management.

Lemma 2. *Optimal conflict management in the all-pay auction induces type independent beliefs in the escalation game, that is, $b_i(1) = b_i(K)$.*

Applying Lemma 2 we define $b_i := b_i(1) = b_i(K)$. The design problem reduces to the question, if the escalation game should be symmetric, $b_1 = b_2$, or asymmetric, $b_1 > b_2$.

The effect of asymmetry on the expected virtual rent is ambiguous. Symmetry guarantees symmetric strategies which in turn imply symmetric payoffs. That is, in the symmetric case we have symmetric intervals $\overline{a}_2^K = \overline{a}_1^K$, and no mass point, $F_1^K(0) = 0$. Both weak players have a continuation payoff after escalation of 0 regardless of their report. The effect of asymmetric beliefs is sketched in figure 3. Starting at $b_1 = b_2$ and gradually increasing b_1 has no effect on the continuation payoff of type-K player 1. The continuation payoff of type-K player 2, however, increases with the degree of asymmetry.

The larger the asymmetry, the more likely the event that player 1 exerts no effort and thus the higher the expected payoff of type-K player 2. Only b_2 determines the strong players' payoff which remains constant in b_1. Consequently, D_1 remains constant while D_2 decreases in b_1. Finally, type independence implies that $b_1 = \rho_2(1)$ and $b_2 = \rho_1(1)$. Thus, increasing b_1 puts higher weight on D_2.

The effect of asymmetry on on-path payoffs in contests is well-known. We graph it in Figure 3(c). The more asymmetric a contest, the less effort players exert. Combining both effects yields a hump-shaped objective, that determines the degree of asymmetry.

Proposition 5. *Suppose the escalation game is the all-pay auction described above. Optimal conflict management induces an on-path belief system in the escalation game with beliefs that are*

- *type independent,* $b_i = b_i(1) = b_i(K)$*, and*
- *have full support,* $b_i \in (0,1)$.

Moreover, beliefs are always asymmetric*,* $b_1 \neq b_2$*. Ignoring public signals is without loss of generality if and only if* $p \leq 1/3$.

Ignoring C_F, we obtain an optimum at $\widehat{\mathcal{B}}$ described by $b_1^* = 1/2 + p/2$ and $b_2^* = 1/2 - p/2$. If the prior $p \in [\underline{r}, 1/3]$ with $\underline{r} := (2(K-1) - \sqrt{8 - 4K + K^2})/(2 + 3K)$, that optimum satisfies C_F. As a consequence, Corollary 1 applies and public signals do not improve. If $p < \underline{r}$, then the reduced form mechanism implementing $\widehat{\mathcal{B}}$ is not feasible and we have to adjust the belief system accordingly. While a closed-form solution may fail to exist, the qualitative statements of Proposition 5 prevail and signals do not improve.

If, however, $p > 1/3$, $\widehat{\mathcal{B}}$ violates incentive compatibility for type-1 player 2. Type-1 player 1's incentive constraint holds with strict inequality. In that case the mechanism implementing $\widehat{\mathcal{B}}$ is enhanced by a public coin flip that determines which of the players takes the role of player 1. With probability 1/2 the mechanism implements and announces $\widehat{\mathcal{B}}$ upon escalation. With the remaining probability she implements the mirror image of that belief system $\breve{\mathcal{B}} = \{\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\beta}}_1\}$.[14]

4.3 Discussion

Proposition 4 and 5 emphasize that conflict management is sensitive to the underlying game form. In particular, we demonstrate how the designer's focus shifts when moving from a game with belief-independent equilibrium actions to one with belief-dependent equilibrium actions. The optimal belief systems in the two settings are in stark contrast. In lotteries, beliefs are *type dependent, symmetric, and partially degenerate.* In the all-pay auction they are *type independent, asymmetric, and never degenerate.*

Economically, the designer's main objective in the lottery case is to determine "easy-to-settle" type profiles. She guarantees full settlement to these profiles. Further, she identifies "impossible-to-settle" profiles that always escalate and intermediate "hard-to-settle cases" that sometimes settle.[15]

In contests, the designer cannot identify such a partitioning of the type space. To the contrary, she implements the same full-support beliefs for every type of a given player. The designer's main objective shifts to deter "obfuscation-driven deviations." By misreporting her type a player gains an information advantage over her opponent. The

[14]Obviously, a *symmetrizing* signal is always feasible and ensures ex-ante symmetry. However, only in case $p > 1/3$ a signal is *necessary* at the optimum. In any case the realized escalation game is played under asymmetric beliefs.

[15]We provide a solution algorithm for a general class of lotteries in appendix C.

deviator correctly forms expectations about her opponent's on-path behavior, but deviations induce faulty updating. Eliminating these incentives is of first-order importance. Incentives to obfuscate vanish if the designer guarantees that learning is independent of a particular report.

Furthermore, belief-dependent actions imply that the expected inefficiencies of the escalation game depend on the belief system. In the all-pay auction, limiting these inefficiencies induces an asymmetric escalation game.

Linking the optimal solution to the properties of the escalation game via Section 3 identifies and quantifies how the designer's main motive shifts with the properties of the game. Our approach highlights that the optimal design of conflict management is shaped by its consequences on the escalation game. If the designer, e.g., an arbitrator, expects players' litigation strategies to be unaffected by beliefs, a sorting mechanism is the conflict management of choice. If, in contrast, she expects players' litigation strategies to vary with beliefs, focus shifts towards deterrence of obfuscation within conflict management.

5 Extensions

In this section we discuss robustness of our findings by proposing several extensions.

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 to hold *at the optimum only.* Arguably the assumption remains strong nonetheless.

Games as outside option necessarily lead to a deviation from the quasi-linearity paradigm used in most mechanism design problems. Assumption 3 recovers some of the properties from quasi linearity and thus ensures tractability. An alternative, perhaps closer, way to restore these properties is to assume *linearity in types.*

Definition 8 (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

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

for any m_i, θ_i, and $\mathcal{B}$.

For a game that is linear in types, the insights of Section 3 remain and the content of Theorem 2 changes only slightly. Depending on binding participation 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.

Even if neither of the conditions hold, the procedure in Section 3 is a necessary first step to solve the problem. If the optimal solution satisfies all global constraints, we have found an optimum. If the optimal solution violates any omitted constraints, we replace the objective with a Lagrangian objective including the global constraint. Results from Theorem 1 and 2 remain under the adjusted objective.[16]

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, all of the properties are either common in the literature ((HC) and (S)) or serve a well-defined purpose (OST). Property (CONV) is special to mechanism design problems in which at least one party can unilaterally enforce a given veto game. 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 on conflicts, for example, those discussed in Section 4, (CONV) is satisfied *for all* possible priors.

We can eliminate assumption (CONV) completely if we assume the designer has access to a Bayesian persuasion device a la Kamenica and Gentzkow (2011) that realizes independently of players participation decision. In addition we have to change the timing slightly: Instead of sequentially ratifying conflict management and then communicating, players do both simultaneously. The designer promises to publish a Bayes' plausible signal over the information she receives in case either of the players rejects the mechanism. This promise (or threat) alone ensures full participation and allows us to employ our techniques even absent of (CONV).[17]

Enlarging the Designer's Signal Space. We restrict the designer's signal space to public messages to focus on the duality of the problems. In the dual problem, the designer has to solve an information design problem using the methods from Bergemann and Morris (2016). 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 the differences in conflict management mechanisms that are embedded in a greater strategic environment beyond the designer's control. Apart from this deviation we 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. In our litigation example we cover what is known as "arbitration". A second important mode known as "mediation" allows players to invoke litigation at any point during conflict management rather than only at the beginning.

[16] A formal treatment of the general Lagrangian is in the supplementary material to this paper.

[17] A general treatment of such signaling mechanisms is discussed in Balzer and Schneider (2017b).

Conceptually, mediation implies that the participation constraints hold ex-post, that is, after players learn the outcome of conflict management. Hörner, Morelli, and Squintani (2015) show for the case of simple lotteries that the optimal arbitration mechanism can also be implemented in a mediation setting. The techniques in Hörner, Morelli, and Squintani (2015) do not generalize to arbitrary games. We propose an alternative extension. Suppose instead of offering the settlement shares publicly, the designer can *privately offer* each party their share. Then, on-path escalation may be triggered via an unacceptable settlement offer and a private recommendation to reject it. Receiving an acceptable offer does hence not imply settlement as the opponent may still reject.

Furthermore, 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 may occur 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. As long as the probability remains positive, however, the designer can react to *any* deviation by implementing the worst possible belief system of the deviating type.[18]

The lower bound on the implementable settlement offer is given by this worst belief. In a companion paper (Balzer and Schneider, 2017a) we show in detail how such a mechanism can be constructed and that in the all-pay auction the settlement ratio of optimal *mediation* is arbitrarily close to that of optimal *arbitration*.

Different Objective. Our choice of objective implies that the designer's preferences are as simple as possible. She focuses exclusively on achieving settlement. In principle, however, the designer may have preferences about the outcome of the escalation game, too. In legal conflicts, the designer may be willing give up some settlement solutions to decrease overall inefficiency in the continuation game.

Adjusting ($\mathrm{P}^{\mathcal{B}}_{max}$) is straightforward. Suppose the designer 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 case of the all-pay auction such preferences would lead 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

$$\frac{\partial \sum_i \mathbb{E}[\Psi_i|\mathcal{B}]}{\partial \mathcal{B}} h_\Psi(\mathcal{B}) + \frac{\partial \sum_i \mathbb{E}[U_i|\mathcal{B}]}{\partial \mathcal{B}} h_U(\mathcal{B}), \qquad \text{with } h_U \geq h_\Psi \geq 0,$$

if the designer cares about reducing inefficiencies conditional on escalation, too.

Implementing Reduced-Form Mechanisms. Our main results are on reduced-form mechanism. While redundant for the escalation games of Section 4, the general implementation conditions, (GI), determine whether the solution is implementable. They serve as an additional constraint to the problem. Two extensions to our model are particularly related to these conditions.

[18] Similar techniques are used in Correia-da-Silva (2017) and Gerardi and Myerson (2007).

The first considers correlated types. If types are correlated, the designer 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. Ex-post settlement outcomes are splitting 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 designer could impose (binding) transfers of any sort in addition 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 then include that players ex-post utility is negative, as they might lose the good *and* pay a transfer to their opponent.[19]

6 Conclusion

We provide a general, yet tractable approach to optimal conflict management within a strategic environment partially beyond the designer’s control. We propose an economically intuitive dual problem to the conflict management 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 effects 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 two escalation games. One in which action choices are invariant to information, and one in which action choices are sensitive to information. We show that a conflict manager’s first-order objective strongly differs. In the first case, the designer prioritizes to partition the set of type profiles. Some cases are solved for sure, some escalate for sure and others escalate with positive probability. In the second case, the designer’s priority is to reduce a disputant’s information advantage from deviation. This information advantage is eliminated if learning from conflict management is independent of the disputant’s behavior within conflict management.

Our results suggest a number of directions for future research. In this paper we focus on conflict management, that is, to find a way to get around a costly resolution of a dispute. However, the property that the design of a mechanism interacts with a greater strategic environment is true beyond conflict management. In fact, the designer in most real-world problems only controls part of the strategic environment, but information obtained while interacting within the mechanism is relevant beyond the mechanism itself. Problems along this line include antitrust measures, financial regulation, and

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

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 a connection exists 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 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 proofs and further discussion on Section 4.

A Belief Management and Proof of Theorem 1

A.1 Proof of Proposition 1

Proof. Full settlement implies pooling. Hence, the sum of players' expected payoffs is constant in their types and can be set to 1. For given p, $V_i(\theta_i, (p, \tilde{p}))$ is decreasing in θ_i. Thus, if $\sum_i V_i(1, (p, \tilde{p})) > 1$ for all $\tilde{p}$, then all pooling solutions violate at least one player's participation constraints. Conversely, the pooling solution $x_i = V_i(1, (p, \tilde{p}))$ implements full settlement. By symmetry a pooling solution exists if and only if an equal split pooling solution exists. □

A.2 Conflict Management and Belief Management

We prove Theorem 1 in steps. The steps correspond to the observations in the main text. Different to the text, we include the designer's ability to provide a public signal in the formal argument. Each realization σ occurs with probability $Pr(\sigma)$. The corresponding realization of the random variable $\mathcal{B}(\Sigma)$ is $\mathcal{B}(\sigma)$.

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, $\beta_i(\theta_{-i}|\theta_i, \sigma)$, 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 9. 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 $\mathcal{B}(\sigma)$. $\square$

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 $Pr(\mathcal{G}) = \sum_{(\theta_1,\theta_2)} p(\theta_1)p(\theta_2)\Big(\sum_\sigma \alpha^\sigma\hat{\gamma}^\sigma(\theta_1,\theta_2)\Big)$. Let $\mathcal{A}$ be the set of all $\{\alpha^\sigma\}_\sigma$, with α^σ such that $\forall(\theta_1,\theta_2)$, $\alpha^\sigma\hat{\gamma}^\sigma(\theta_1,\theta_2) \leq 1$ and $\hat{\gamma}(\theta_1,\theta_2) = \sum_\sigma \alpha^\sigma\hat{\gamma}^\sigma(\theta_1,\theta_2) \leq 1$. $\mathcal{A}$ 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: Linearity of constraints 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. $\square$

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 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 $\Theta_i^I(\theta_i) \cap \{\Theta_i^{PC} \cup \Theta_i^0 \cup \{\Theta_i^I(k)\}_{k\notin\widehat{\Theta}_i}\} = \emptyset$. $\square$

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. $\square$

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

$$\begin{aligned}&\sum_{\theta_i} p(\theta_i) z_i(\theta_i) = \\ &\quad -\sum_\sigma H_i(\gamma^\sigma) + \sum_{\theta_i \in \Theta_i^{PC}} p(\theta_i) V_i(\theta_i,(p,\rho^V)) + \sum_{\theta_i \in \Theta_i^{IC}} p(\theta_i) \sum_{k \in \Theta^I(\theta_i)} V_i(k,(p,\rho^V))\end{aligned} \tag{7}$$

Let $H_i(\{\gamma^\sigma\}_\sigma) := \sum_\sigma H_i(\gamma^\sigma)$. Further let $P_i(\Theta_i^0) := \sum_{\theta_i \in \Theta_i^0} p(\theta_i)$. Straight-forward 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. A conflict management outcome is only feasible if the ex-ante expected settlement values are weakly lower than the probability of settlement, (2). That is, $\sum_i \sum_{\theta_i} p(\theta_i) z_i(\theta_i) \leq 1 - Pr(\mathcal{G})$, 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 (2). By step 3b we can rewrite (2) as

$$\sum_i v_i(V_i(\Theta,(p,\rho^V))) - 1 \leq \sum_\sigma \left(\sum_i \Big(P_i(\Theta_i^0)(\alpha^\sigma - 1)H_i(\gamma^\sigma) + H_i(\gamma^\sigma)\Big) - Pr(\mathcal{G},\sigma) \right) \tag{2'}$$

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

$$\underbrace{\sum_\sigma \left(\sum_i \Big(P_i(\Theta_i^0)(\alpha^\sigma - 1)H_i(\gamma^\sigma) + H_i(\gamma^\sigma)\Big) - \alpha^\sigma \sum_{(\theta_1,\theta_2)\in\Theta} p(\theta_1)p(\theta_2)\gamma^\sigma(\theta_1,\theta_2) \right)}_{=:h(\{\alpha^\sigma\gamma^\sigma\}_\sigma)}.$$

Moreover, using the definition of H_i it follows that h is linear in α, since $y_i(\theta_i;\theta_i)$ and $Pr(\mathcal{G})$ 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 (2') 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 The Dual Problem and Proof of Theorem 2

We construct a solution algorithm to solve for $\mathcal{CM}$. 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 9 in the supplementary material provides the corresponding formal argument.

B.1 Proof of Proposition 2

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

B.2 Proof of Lemma 1

Proof. The MDR property implies that local downward incentive compatibility is sufficient for global downward incentive compatibility.[20] Now, assume there is a type θ_i for which both incentive constraints are redundant. Then, it is possible to reduce $z_i(\theta_i)$ at no cost for the designer until either an incentive constraint binds, i.e., $\theta_i \in \Theta_i^{PC}$, the participation constraint starts to bind, i.e., $\theta_i \in \Theta_i^{PC}$, or $z_i = 0$, i.e., $\theta_i \in \Theta_i^0$.

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 and players are symmetric, and so does the combination. Under the alternative mechanism, no participation constraint is binding. A contradiction.

To see that exactly 1's participation constraint is binding and that $z_i(k) \geq 0$ does not bind, consider the designer's resource constraint. Focus on the formulation (2') in the proof of Theorem 1, step 4. Assume by contradiction that the set of types with binding participation constraint $\Theta_i^{PC} \neq \{1\}$. An upper bound on LHS of (2') 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, $\Theta_i^{PC} = \{1\}$, it is without loss to focus on mechanisms in which the downward local incentive compatibility binds: Suppose $\theta_i > 1$'s downward IC is redundant. The designer can reduce $z_i(\theta_i)$ (and potentially burn the share) without violating any other constraint. □

B.3 The Lagrangian Problem

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

[20] 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 appendix D in the supplementary material.

Lemma 3. *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 $Pr(\mathcal{G})$ a function from choices to escalation probabilities. Define $\mathrm{w}(t) := inf\{Pr(\mathcal{G})|cs = (\gamma, z, \Sigma) \in CS, G(cs) \leq t\}$. 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 $cs_j = (\gamma[j], z[j], \Sigma[j])$ describe the optimal solution, such that $Pr(\mathcal{G})(cs_j) = \mathrm{w}(t_j)$. 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

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

A contradiction. $\square$

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

Definition 10 (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 $\nu^i_{\theta,\theta'}$ for the incentive compatibility 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 2. Finally, μ_{θ_1,θ_2} is the multiplier on the consistency constraint. We divide all multipliers by δ and obtain the set $\{\nu^{\tilde{i}}_{\theta,\theta'}, \tilde{\lambda}^i_\theta, 1, \tilde{\zeta}_{\theta^i}, \tilde{\eta}_Q, \tilde{\mu}_{\theta_1,\theta_2}\}$. Let $\tilde{e}_i(\theta) = p(\theta) \sum_{Q|\theta\in Q} \eta_Q$. 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, $\mathsf{m}^{\mathsf{i}}_\theta := p(\theta) + \tilde{e}^i_\theta - \tilde{\zeta}^i_\theta$, $\mathsf{M}^{\mathsf{i}}(\theta) := \widetilde{\Lambda}^i(\theta) - \sum_{k=1}^{k=\theta} p(k) - \widetilde{E}^i(\theta) + Z^i(\theta)$, $\nu^i(\theta) := \sum_{k=1}^K |(\sum_{l=1}^K \tilde{v}^i_{k+1,l})| - \tilde{v}^i_{\theta,\theta+1}$.

We state the transformed Lagrangian objective as a corollary to the more general solution discussed in the supplementary material appendix E.

Corollary 2. *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*

$$
\begin{aligned}
\widehat{\mathcal{L}}(\mathcal{B}(\sigma)) :=& \mathcal{T}(\mathcal{B}(\sigma)) + \sum_i \Bigg[\sum_{\theta=1}^{K} \rho_i(\theta|\sigma)(\frac{\mathsf{m}^{\mathsf{i}}_\theta}{p(\theta)})U_i(\theta;\theta,\mathcal{B}(\sigma)) \\
&+ \sum_{\theta=1}^{K-1} \frac{\mathsf{M}^{\mathsf{i}}(\theta) - \tilde{\nu}^i_{\theta,\theta+1}}{p(\theta)} \rho_i(\theta|\sigma)(U_i(\theta;\theta,\mathcal{B}(\sigma)) - U_i(\theta;\theta{+}1,\mathcal{B}(\sigma))) \\
&- \sum_{k=1}^{\theta-1} \frac{\mathsf{M}^{\mathsf{i}}(\theta) + \nu^i_{k,\theta} - \nu^i(\theta)}{p(\theta)} \rho_i(\theta|\sigma)[U_i(\theta;k,\mathcal{B}(\sigma)) {-} U_i(\theta;\theta,\mathcal{B}(\sigma))] \Bigg],
\end{aligned} \tag{8}
$$

with

$$
\mathcal{T}(\mathcal{B}(\sigma)) := \sum_{Q \in Q^2} \sum_{(\theta_1,\theta_2)\in\tilde{Q}} [\rho(\theta_1|\sigma)\beta_1(\theta_2|\theta_1,\sigma)]\tilde{\eta}_Q - \sum_{\theta_1\times\theta_2} \frac{\rho_1(\theta_1|\sigma)\beta_1(\theta_2|\theta_1,\sigma)}{p(\theta_1)p(\theta_2)} \tilde{\mu}_{\theta_1,\theta_2}. \tag{9}
$$

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* (2) *binds, hence* $\delta > 0$*.*
- *If the optimal reduced form mechanism is implementable, then the constraints from Proposition 2 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* $\tilde{v}^i_{\theta,k} = 0$ *for all* $k \geq \theta$*.*
- *Local downward incentive constraints are sufficient for global incentive constraints.*

Results follows from algebraic manipulation of the initial Lagrangian objective using Lemma 1 to identify binding constraints. Manipulations proceed alongside the discussion of Proposition 3. A full description is in appendix E of the supplementary material.

B.4 Proof of Proposition 3

Proof. With access to signal 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 2. Take the first two terms of the Lagrangian in Corollary 2 as the objective due to the binding constraints from Lemma 1, 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 2. Consequently, $\frac{\mathsf{m}^{\mathsf{i}}_\theta}{p(\theta)} = 1$ and $(\mathsf{M}^{\mathsf{i}}(\theta) {-} \tilde{\nu}^i_{\theta,\theta+1})/p(\theta) = w(\theta)$. The last term of (8) boils down to an expression that consists of local downward incentive constraints. The signaling term (9) is implied by consistency and the border constraints, completing the proof. ☐

B.5 Proof of Corollary 1

Proof. The solution to the optimization problem, $\mathcal{B}^*$, maximizes (8). By hypothesis, $\mathcal{B}^*$ is in the set of least constraint solutions. Thus the last two terms of equation (8) 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. □

B.6 Proof of Theorem 2

Proof. The problem collapses to ($\mathrm{P}^{\mathcal{B}}_{max}$) if the optimal signal is degenerate and Corollary 1 applies. Suppose we are at an optimum with a non-degenerate signal. Assume that instead of the continuation game $\mathcal{G}$, an alternative continuation game $\widehat{\mathcal{G}}$ is played. $\widehat{\mathcal{G}}$ differs from $\mathcal{G}$ 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 $\mathcal{G}$ under updated beliefs. The continuation payoff of $\hat{\mathcal{G}}$ 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, $\mathbb{E}[\widehat{U}|\mathcal{B}(\Sigma)] := \sum_i \sum_\sigma Pr(\sigma)\mathbb{E}[U_i(\theta;\theta,\mathcal{B}(\sigma))|\mathcal{B}(\sigma)]$. Any mean $\overline{\mathcal{B}}$ may have many possible lotteries that support it and are feasible. We select the maximum for each. Hence, $\overline{\mathcal{B}}$ that maximizes $\mathbb{E}[\widehat{\Psi}|\mathcal{B}(\Sigma)] + \mathbb{E}[\widehat{U}|\mathcal{B}(\Sigma)]$ also solves (8). □

C Proofs and Additional Material to Section 4

C.1 Conflict Management in Monotone Lotteries

In this section we provide a general solution algorithm for *monotone lottery games* with an arbitrary number of types. The solution nests that in the main text.[21]

Definition 11 (Lottery Game)**.** $\mathcal{G}$ is a lottery game, if $u(\theta_i,\theta_{-i},\mathcal{B})$ is constant in $\mathcal{B}$.

Consistent with the main text we suppress the argument $\mathcal{B}$ in u.

Definition 12 (Monotone Lottery Game)**.** A lottery game is monotone if $u(\theta_i,\theta_{-i}) - u(\theta_i+1,\theta_{-i})$ is weakly decreasing in θ_i and θ_{-i} and the prior p induces a weakly decreasing inverse hazard rate $w(\theta_i)$.

The lottery feature entails that expected continuation payoffs are linear in any $\beta_i(\cdot|m_i)$. Therefore, the problem is reminiscent of a standard mechanism design problem with interdependent values. The solution is thus given by a linear program.

We define the ex-ante probability that the type profile (θ,k) will face each other in litigation as $\rho(\theta,k) := \rho_1(\theta)\beta_1(k|\theta)$ with $\sum_{(\theta,k)} \rho(\theta,k) = 1$. Naturally, $\rho(1,1) = 1$ implies $\beta(1|1) = 1$ and 0 for all other beliefs. Substituting for $\beta_i(\theta_i|\theta_{-i})$ and rearrange yields the function $\widetilde{VV}(\theta_1,\theta_2)$, which is independent of any $\rho(\cdot,\cdot)$. The objective of

[21] A matlab program implementing the algorithm is available from the authors.

($\mathrm{P}^{\mathcal{B}}_{max}$) becomes

$$\Xi(\rho(\cdot,\cdot)) := 2u(K,K) + \sum_{\Theta^2\backslash(K,K)} \rho(\theta_1,\theta_2)\widetilde{VV}(\theta_1,\theta_2). \tag{10}$$

We define the condition

$$(2V(1,(p,\underline{\rho}^V)) - 1)\rho(1,1) \leq (p(1))^2 \left(\Xi(\rho(\cdot,\cdot)) - 1\right). \tag{11}$$

Definition 13 (Top-Down Algorithm)**.** Denote the set of type pairs (k,θ) such that $\rho(k,\theta) > 0$ by Θ^2_+. Start with $\Theta^2_+ = \emptyset$.

1. Set $\rho(1,1) = 1$ and check whether condition (11) is satisfied. If it is satisfied, then terminate. Otherwise continue at 2.
2. Identify the set $\Theta^2_N = \{(\theta_1,\theta_2)|(\theta_1,\theta_2) = \arg\max_{\Theta^2\backslash\Theta^2_+} \widetilde{VV}(\theta_1,\theta_2)\}$.
   (a) Set $\rho(1,1)$ to the solution of

   $$\sum_{\substack{(\theta_1,\theta_2)\in \\ \Theta^2_+\cup\Theta^2_N}} \frac{p(\theta_1)p(\theta_2)}{(p(1))^2}\rho(1,1) = 1. \tag{12}$$

   (b) Replace $\rho(\theta_1,\theta_2) = \frac{p(\theta_1)p(\theta_2)}{(p(1))^2}\rho(1,1)\ \forall(\theta_1,\theta_2) \in \Theta^2_+ \cup \Theta^2_N$.
   (c) Check whether condition (11) is satisfied. If it is satisfied, decrease all ρ for the set Θ^2_N at the expense $\rho(1,1)$ keeping the relation of 2(b) for $\Theta^2_+ \setminus \Theta^2_N$ until the condition holds with equality. Then, terminate. If (11) is violated, repeat step 2.

Proposition 6. *Suppose the escalation game is a monotone lottery. Then optimal conflict management is the solution to the top-down algorithm.*

Proof. In the supplementary material, appendix H, as it is mainly technical. ☐

C.2 Conflict Management in the All-Pay Auction

C.2.1 Closed-Form Expression of Effort Distributions

We assume without loss of generality that $b_1(1) \geq b_2(1)$. We identify players via their marginal cost θ_i. We start by deriving equilibrium strategies.

Players' 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. Player θ_i's support includes a if and only if a maximizes

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

Referring to Figure 1 on page 21, the support of players' strategies can be partioned into $I_1 = (0, \overline{a}^K_2]$, $I_2 = (\overline{a}^K_2, \overline{a}^K_1]$ and $I_3 = (\overline{a}^K_1, \overline{a}^1_1]$. 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. Player θ_i mixes such that the opponent's first-order condition holds on the joint support. The densities are

$$
\begin{aligned}
f_2^1(a) &= \mathbb{1}_{\in I_2}\frac{K}{b_1(K)} + \mathbb{1}_{\in I_3}\frac{1}{b_1(1)}, & f_2^K(a) &= \mathbb{1}_{\in I_1}\frac{K}{1-b_1(K)}, \\
f_1^1(a) &= \mathbb{1}_{\in I_3}\frac{1}{b_2(1)}, & f_1^K(a) &= \mathbb{1}_{\in I_1}\frac{K}{1-b_2(K)} + \mathbb{1}_{\in I_2}\frac{1}{1-b_2(1)}.
\end{aligned}
$$

This leads to the following cumulative density functions:

$$
\begin{aligned}
F_2^1(a) &= \mathbb{1}_{\in I_2} a\frac{K}{b_1(K)} + \mathbb{1}_{\in I_3}\left(\frac{a}{b_1(1)} + F_2^1(\overline{a}_2^K)\right) + \mathbb{1}_{>I_3}, \\
F_2^K(a) &= \mathbb{1}_{\in I_1} a\frac{K}{1-b_1(K)} + \mathbb{1}_{>I_1}, \\
F_1^1(a) &= \mathbb{1}_{\in I_3}\frac{a}{b_2(1)} + \mathbb{1}_{>I_3}, \\
F_1^K(a) &= \mathbb{1}_{\in I_1}\left(a\frac{K}{1-b_2(K)} + F_1^K(0)\right) + \mathbb{1}_{\in I_2}\left(\frac{a}{1-b_2(1)} + F_2^K(\overline{a}_2^K)\right) + \mathbb{1}_{>I_2}.
\end{aligned}
$$

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

1. $\overline{a}_2^K$ is determined using $F_2^K(\overline{a}_2^K) = 1$, i.e., $\overline{a}_2^K f_2^K(a) = 1$ with $a \in I_1$. Substituting yields
$$\overline{a}_2^K = \frac{1-b_1(K)}{K}.$$
2. For any $\overline{a}_1^K$, $\overline{a}_1^1$ is determined using $F_1^1(\overline{a}_1^1) = 1$, i.e., $\left(\overline{a}_1^1 - \overline{a}_1^K\right) f_1^1(a) = 1$ with $a \in I_3$. Substituting yields
$$\overline{a}_1^1 = \overline{a}_1^K + b_2(1).$$
3. $\overline{a}_1^K$ is determined by $F_2^1(\overline{a}_1^K) = 1$. That is, $\left(\overline{a}_1^K - \overline{a}_2^K\right) f_2^1(a) + \left(\overline{a}_1^1 - \overline{a}_1^K\right) f_2^1(a') = 1$ with $a \in I_2, a' \in I_3$. Substituting yields
$$\overline{a}_1^K = \overline{a}_2^K + \left(1 - \frac{b_2(1)}{b_1(1)}\right)\frac{b_1(K)}{K}.$$
4. $F_1^K(0)$ is determined by the condition $F_1^K(\overline{a}_1^K) = 1$, i.e., $F_1^K(0) = 1 - \overline{a}_2^K f_1^K(a) - \left(\overline{a}_1^K - \overline{a}_2^K\right) f_1^K(a')$ with $a \in I_1, a' \in I_2$. Substituting yields
$$F_1^K(0) = 1 - \frac{1-b_1(K)}{1-b_2(K)} - \left(1 - \frac{b_2(1)}{b_1(1)}\right)\frac{b_1(K)}{1-b_2(1)}\frac{1}{K}.$$

Change of Variables. To simplify the argument in the proof of Proposition 5 we express $\mathcal{B}$ entirely using ρ and $b_1(1)$. The probabilities ρ are the solution to the system of linear equations $\rho_i(\theta_i) = \sum_{k=1}^K \beta_{-i}(\theta_i|k)\rho_{-i}^{\mathcal{G}}(k)$. We abuse notation and define $\rho_i :=$ $\rho_i(1)$. Given ρ_1, ρ_2 we describe any $b_i(m) \in \mathcal{B}$ as a linear function of $b_1(1)$ using Bayes' rule. That is,

$$b_1(K) = \frac{\rho_2 - \rho_1 b_1(1)}{1-\rho_1}, \quad b_2(K) = \frac{\rho_1}{1-\rho_2}(1-b_1(1)), \quad \text{and } b_2(1) = \frac{\rho_1}{\rho_2} b_1(1).$$

Closed Form Expressions relative to $b_1(1)$. Any fraction $\beta(\theta_1|\theta_2)/\beta(\theta_2|\theta_1)$ depends only on ρ, and we simplify further

$$\begin{aligned} \overline{a}_2^K &= \frac{1-\rho_2-\rho_1}{K(1-\rho_1)} + \frac{\rho_1}{K(1-\rho_1)} b_1(1), & F_1(0) &= (1-b_2(m)\left(\frac{(\rho_2-\rho_1)}{(1-\rho_1)}\frac{(K-1)}{K}\right), \\ F_2(\overline{a}_2^K|m) &= (1-b_1(m)), & F_1(\overline{a}_2^K|m) &= (1-b_2(m))\left(1-\frac{(\rho_2-\rho_1)}{(1-\rho_1)}\frac{1}{K}\right). \end{aligned} \tag{13}$$

C.2.2 Proof of Proposition 5

Structure of the proof. We prove Lemma 2 together with Proposition 5. We use a guess and verify approach to prove the statements jointly. A constructive proof is possible, but notationally intense. We omit showing that the escalation game has a unique and monotonic equilibrium at the optimum and the case for small priors $p < \underline{r}$. Both aspects are straightforward to verify. A more constructive version including these missing steps is in a companion paper Balzer and Schneider (2017a). We start by guessing that the conditions in Corollary 1 hold and no additional constraint binds.

Proof. ***Part A (Piece-wise Linearity).*** $b_1(1) \geq b_2(1)$ implies $\rho_2 > \rho_1$. Take any ρ_i that satisfies this condition. Expressions (13) imply that a player's winning probability, $F_i(\overline{a}_2^K|m)$, is linear in $b_1(1)$, since $(1-b_i(m))$ is linear in $b_1(1)$. Since $\overline{a}_2^K$ is linear in $b_1(1)$, too, so are the payoffs, (5). The virtual rent, (6), is piecewise linear with a kink at $b_1(1) = \rho_2$.

Thus, the objective is (piece-wise) linear. Multiplying with p for readability yields

$$\begin{aligned} \Xi(b_1(1)) &:= p\left(\sum_i \mathbb{E}[\Psi_i|\mathcal{B}] + \mathbb{E}[U_i|\mathcal{B}]\right) \\ &= (\rho_1+\rho_2)U_1(1;1,\mathcal{B}) - (1-p)\sum_i \rho_i U_i(1;K,\mathcal{B}) + p(1-\rho_2)U_2(K;K,\mathcal{B}). \end{aligned} \tag{14}$$

Part B (Optimality).

Step 1: Type-independence. Linearity immediately implies that the optimal $b_1(1)$ 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

$$\underline{b} = \frac{\rho_1}{K(1-\rho_2)+\rho_2}, \qquad \overline{b} = \frac{(K-1)(1-\rho_1)+\rho_2}{K(1-\rho_1)+\rho_1}, \qquad b^* = \rho_2.$$

We guess that the optimum is at $b^* = \rho_2$ taking Lemma 2 for granted and proceed.

Step 2: Type distribution. Given Lemma 2 we now determine the optimal ρ_i. Using the RHS of (14) and substituting $b_1(1) = \rho_2$ yields a quadratic objective in both ρ_is. Moreover, the first-order conditions are independent of each other. The unique solution is $(\rho_1, \rho_2) = ((1-p)/2, (1+p)/2)$.[22] Second-order conditions are satisfied at the desired point and we can conclude that a local optimum exist in case we face a least constraint 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, (2), with equality.

Step 3: Upward Incentive Constraints and Potential for signals. Downward incentive constraints are satisfied 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 condition becomes $\gamma_i(1)U_i(1;1,\widehat{\mathcal{B}}) - \gamma_i(K)U_i(K;1,\widehat{\mathcal{B}}) \geq \gamma_i(1)U_i(1;K,\widehat{\mathcal{B}}) - \gamma_i(K)U_i(K;K,\widehat{\mathcal{B}})$. Using type-independence we have that $U_i(1;k,\widehat{\mathcal{B}}) = U_i(K;k,\widehat{\mathcal{B}})$.

Incentive compatibility is thus satisfied if

$$\gamma_i(1) \geq \gamma_i(K) \Leftrightarrow \rho_i \geq p. \tag{15}$$

This always holds for player 1_2, but not for player 1_1 if $p > 1/3$. Now consider the following mechanism with public signals. There are two realizations, σ_1 and σ_2, both equally likely. If σ_1 realizes the mechanism proceeds as above, if σ_2 realizes, the mechanism flips players' identities. By ex-ante symmetry, the value of the problem remains constant and condition (15) holds by Assumption 1 as it becomes

$$(\gamma_i^{\sigma_1}(1) + \gamma_i^{\sigma_2}(1)) \geq (\gamma_i^{\sigma_1}(K) + \gamma_i^{\sigma_2}(K)) \Leftrightarrow \frac{1}{2} \geq p.$$

Step 4: Verifying local optimality. We now verify that type independence yields a local optimum. Assume to the contrary that $b_1(1) < \rho_2$ at the optimum. Substituting the claimed optimum into the objective we observe that

$$\begin{aligned} \Xi(b_1(1))|_{b_1(1)<\rho_2} = {} & F_1^K(\overline{a}_2^K)\left((1-b_1(1))\,\rho_2 + p(1-\rho_2)\right) \\ & - (1-b_2(1))\left((1-p)\rho_2 F_1^K(\overline{a}_2^K)\right) \\ & + \underbrace{\overline{a}_2^K}_{=(1-b_1(K))/K}\left((\rho_1+\rho_2)(K-1) - (1+\rho_2)Kp\right). \end{aligned} \tag{16}$$

The derivative changes sign at $b_1(1) = \rho_2$ since the derivative of the solution evaluated

[22] By continuity of the objective the same holds true if we take the objective given $b_1(1) \geq \rho_i$ instead.

at the optimal point $\rho^* = (\rho_1 = (1-p)/2, \rho_2 = (1+p)/2)$ reads

$$\frac{\partial \Xi(b_1(1))}{\partial b_1(1)}|_{\rho^*} = \begin{cases} \frac{K(1-(p)^2)-(1-(p)^2)}{K(1+p)} & \text{if } b_1(1) < \rho_2 \\ -\frac{K(1-(p)^2)-(1-(p)^2)}{K(1+p)} & \text{if } b_1(1) > \rho_2 \\ \text{undefined} & \text{if } b_1(1) = \rho_2. \end{cases}$$

Step 5: Global Optimality. For global optimality we have to verify that the optimal solution given $b_1(1) \in \{\underline{b}, \bar{b}\}$ is worse than the optimum calculated here. Plugging into Ξ and solving yields the desired result.

Step 6: Implementability. Finally, we have to verify that the reduced-form mechanism is implementable. For that, we plug our solution into the constraints from Proposition 2 and obtain that these are satisfied. □

References

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 (2017a). "Optimal Alternative Dispute Resolution". *mimeo.*

— (2017b). "Persuading to Participate: Mechanism Design with Informational Punishment". *mimeo.*

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.

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.*

Fieseler, K., T. Kittsteiner, and B. Moldovanu (2003). "Partnerships, lemons, and efficient trade". *Journal of Economic Theory* 113, pp. 223–234.

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

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

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.*

Meirowitz, A., M. Morelli, K. W. Ramsay, and F. Squintani (2017). “Dispute Resolution Institutions and Strategic Militarization”. *mimeo.*

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.

Myerson, R. B. and M. A. Satterthwaite (1983). “Efficient mechanisms for bilateral trading”. *Journal of Economic Theory* 29, pp. 265–281.

Philippon, T. and V. Skreta (2012). “Optimal Interventions in Markets with Adverse Selection”. *American Economic Review* 102, pp. 1–28.

Posner, R. A. (1973). “An economic approach to legal procedure and judicial administration”. *The Journal of Legal Studies*, pp. 399–458.

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.

Tirole, J. (2012). “Overcoming Adverse Selection: How Public Intervention Can Restore Market Functioning”. *American Economic Review* 102, pp. 29–59.

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

Zheng, C. (2017). “A necessary and sufficient condition for Peace”. *mimeo.*