Persuading to Participate: Mechanism Design with Informational Punishment

Benjamin Balzer* Johannes Schneider†

May 27, 2017

Preliminary and Incomplete - Do not circulate!

Abstract

We use Bayesian persuasion techniques to relax the participation constraints in mechanisms that are offered as an alternative to a known default game. We show that a mechanism can punish a deviator by releasing *information about the complying players* and thereby persuade the deviator to participate. Based on this observation we provide a simple augmentation to the classical mechanism design problem that ensures full-participation independent of the underlying default game or the prior information structure, and that reduces each players participation constraint to its convex hull over beliefs about her opponents. The only requirement on the designer for informational punishment is that she can commit to send a public signal in case the mechanism is vetoed by one or more players.

1 Introduction

In this paper, we study how Bayesian persuasion can help a mechanism designer to reduces the players participation constraint. In particular, we study how an ex-ante uninformed designer can threat potential deviator with partially releasing the information obtained from participating parties. We show that in the general framework of veto-constraint mechanism this threat induces full-participation at the optimum.

A common theme in economics is to design mechanisms, institutions or game-forms that outperform a status quo defined as the result of the strategic actions without the planner's intervention. Often the greater institutional framework restricts interventions to voluntary mechanisms, that is mechanisms in which each party involved can freely decide whether to

*UT Sydney, benjamin.balzer@uts.edu.au
†Carlos III de Madrid, jschneid@eco.uc3m.es

participate. Voluntary participation becomes in particular relevant, if players interact in a given non-cooperative *default game* should they fail to coordinate on the mechanism. Even more so, if the right to enforce the default game is protected by natural or constitutional rights, each player is given a veto power concerning the mechanism causing an externality on all other players: Each *single* player can veto the mechanism altogether and force the play of the default game on *all* player.

Veto-mechanisms operating *in the shadow* of a default game abound. Examples include, but are not limited to, default games in which players compete for a prize in a contest such as litigation, crowdsourcing, patent races, strikes or quantity competition. From the point of view of the players, these mechanisms are inefficiently costly. Thus, there is an incentive to reach a collusive agreement such as an out-of-court settlement, an industry standard, a collective agreement or a collusive production plan. However, any player is free to participate in the collusion mechanism and cannot be excluded from participating in the default game. Often, a public veto of one of the players to the collusive mechanism leads to the breakdown of the entire mechanism either because full participation is needed by design as in the case of out-of-court settlements or because the non-participant acts as a whistleblower as in the case of collusive production plans.

In this paper we study a mechanism design problem in such an environment. Each player has the option to publicly veto the mechanism and enforce the play of a Bayesian default game. We augment the traditional tools of the designer by the possibility to provide players with an informative signal even if the mechanism has been rejected. That is, while we assume that players can commit not to participate in the mechanism itself they cannot commit to ignore any information available to them. We show that the provision of an informative signal serves as a punishment to a potential deviator and thus relaxes the players' participation constraint. In particular, while the possibility of a public veto without such informational punishment implies that an optimal mechanism might be rejected on the equilibrium-path, our main result, Proposition 1, states that full-participation mechanisms are always optimal if the designer has access to informational punishment.

An attractive feature of informational punishment is it separates the signalling effect of a veto from incentive compatibility. Informational punishment itself exclusively effects the players' participation constraint but has neither a direct effect on the (expected) outcome of the mechanism nor incentive compatibility. The reason is that informational punishment only operates off the equilibrium path. Using methods from the literature on Bayesian persuasion we obtain a simple expression for the participation constraint if informational punishment is available. Having established the relaxed participation constraints allows to solve the remaining mechanism design problem using standard and well-documented techniques. Informational punishment is thus a simple and straight-forward approach to

incorporate veto mechanisms into the existing mechanism design framework.

Our approach is constructive. The designer ex-ante commits both to conduct an experiment on the information it obtains from the participating players and to communicate the result of that experiment. If players cannot commit to ignore the information, every player updates her belief and subsequent behaviour accordingly once the information is revealed. In principle the designer can freely pick any posterior belief as long as the set of all posteriors are Bayes' plausible in the sense that they combine to the prior. The experiment then splits the prior information structure such that deviator's expected payoff before receiving the information is the convex combination of the equilibrium play under each realizations. The deviator's expected payoff are described by the *value of vetoing*, a function that describes a player's expected payoff given any information structure.

Despite operating only off the equilibrium path, informational punishment reduces the players participation constraints by exploiting non-convexities in a player's value of vetoing. Our findings apply to a wide range of mechanisms including settings in which the players can –upon acceptance– fully commit to the mechanism and setting in which the mechanism only works as a coordination device that recommends players an action in the default game.

There are several interpretations of a credible veto consistent with our general framework. A player may refuse to participate in any information exchange with the mechanism, for example, by verifying to have been absent at a scheduled meeting, or by refusing to sign any legally binding documents. In any of these cases ex-ante commitment to ignore information is considerably harder. In particular, if the information is sent by a mechanism that itself has the power to commit to conduct her experiment.

Consider the example of the legal dispute from above. It is considerably simple to reject any settlement attempt by refusing to sign any proposed agreement. However, when entering the formal litigation process after failed settlement negotiations, it is interim optimal for any party to use any information provided when preparing her case. Thus, any ex-ante statement on the handling of new information is not sequentially rational and can therefore be considered as an empty threat by the opponent. Thus, any information the mediator releases about the case prior to litigation influences players strategies and thereby expected payoffs. Expecting such a release in turn influences their participation decision in the first place.

1.1 Related Literature

We contribute to the small but growing literature on mechanism design with endogenous outside options. Our setup fits into that of Compte and Jehiel (2009) although, different to them, we allow for an outside option that is not only type, but also belief dependent. This leads to an informational externality of the choice of mechanism as in Jehiel and

Moldovanu (2001) and Bester and Wärneryd (2006). The main difference to these models is that the Bayesian default game can generate an outside option that is non-convex in beliefs in contrast to the linear outside options assumption these papers make.

Celik and Peters (2011) study optimal collusion with a non-linear outside option. They provide an example within their framework in which a full-participation mechanism is not optimal and the revelation principle fails. Similar to us, they allow for public vetoing of a mechanism, but do not allow for any communication between the mechanism and a potential deviator. We show that once we allow the mechanism to publicly reveal information, an optimal solution with full-participation always exists.

A second concern in Celik and Peters (2011) is that veto mechanisms may violate the principle of inscrutability in an environment in which the designer is an informed principle (Maskin and Tirole, 1990, 1992). Allowing for informational punishment eliminates this issue and restores the principle of inscrutability. Therefore, informational punishment allows to extend the classical toolbox of mechanism design to a general Bayesian environment guaranteeing procedural simplicity even with potentially complicated outside options.

An alternative approach to a similar problem is discussed in Gerardi and Myerson (2007) and Correia-da-Silva (2017) who, too, consider unilaterally veto rights. Both propose a trembling device triggering spurious veto to overcome adverse effects due to off-path beliefs. In our setup, a veto is publicly verifiable, and a trembling device therefore cannot alter off-path beliefs or prevent a veto by a player. Informational punishment focuses instead on the threat of strategic information release in situations that almost surely do not occur. Despite its irrelevance for on-path outcomes, the mere promise of such information release disciplines players to participate.

Our approach offers a rich set of applications. Results apply whenever a designer can choose from a set in the class of veto-constraint mechanisms augmented with public or private signals The richness of the mechanisms' outcome space is only constraint by considerably weak assumptions. Informational punishment is possible in a wide range mechanisms from pure coordination, where an outcome is a recommendation about the player's action in a default game, to traditional mechanism design, where an outcome is a physical vector of goods (often a transfers and a consumption good) and the default-game is payoff irrelevant whenever players decide to participate in the mechanism.

Economically these types of mechanisms can be found in collusion in auctions (McAfee and McMillan, 1992; Balzer, 2016), vertical relations (Hart and Tirole, 1990; McAfee and Schwartz, 1994), innovation (Lerner and Tirole, 2004, 2015), bail-outs (Bolton and Skeel, 2010; Philippon and Skreta, 2012; Tirole, 2012), or dispute resolution (Hörner, Morelli, and Squintani, 2015; Balzer and Schneider, 2017a,b; Zheng, 2017). We offer a simple, yet powerful extension to the canonical setup increasing players' incentives to participate.

Notably, the designer does not expect to pay the cost of punishing a potential deviator as informational punishment becomes relevant only off the equilibrium path.

Methodologically, informational punishment relies on a particular version of information design (Bergemann and Morris, 2016b), known as Bayesian persuasion (Kamenica and Gentzkow, 2011). In our model persuasion is more subtle, though. By threatening players to run an experiment in case of non-compliant behaviour, the designer persuades players to participate in the proposed mechanism. The threat of performing such an experiment convexifies and thus reduces the players participation constraint given their prior in the sense of Aumann and Maschler (1995). We show that the pure existences of such modes of communication increases the designer's options even if on-path communication never happens.

The remainder of this paper is structured as follows: In Section 2 we use an all-pay auction to convey the basic intuition behind informational punishment. In Section 3 we generalise these insights and provide the main result. We discuss our findings in light of a variety of applications in Section 4 and conclude in Section 5.

2 Introductory Example: All-Pay-Auction with Reserve Price

2.1 Underlying Default Game

Two players engage in an all-pay contest over a pie of value 1. Both players simultaneously decide on a score $b_i \geq 0$ and the player with the highest score wins the entire pie. Ties are broken at random. In addition, there is a minimum score $r > 0$. If any of the contestants scores below r, she does not win the pie with probability 1.[1] Obtaining a score is costly and marginal cost are constant with $c_i \in \{1, \kappa\}$ with $\kappa > 1$. The marginal cost are independently drawn according to a distribution characterized by the probability p_i that player i has marginal cost $c_i = 1$. A player's realization of the cost draw is her private information. All other aspects of the game are commonly known. Without loss of generality we restrict attention to cases in which $p_1 \geq p_2$, that it is more likely that player 1 has low marginal cost of effort.

Players' strategies and thus their equilibrium payoffs depend on the relation between the two ex-ante distributions p_1, p_2, the cost disadvantage of the high-cost player κ, and the reserve price r. We say that $I = (p_1, p_2)$ is the commonly known information structure at the point where players make their decisions. Given $p_1 \geq p_2$ the set of possible information

[1] Examples of these types of all-pay auctions are for example crowd sourcing campaigns such as the inducement prize contests such as those designed by the *X prize foundation*. Players have to achieve a certain minimum goal *and* beat there opponents to win the prize. An alternative interpretation of the model may be the litigation process when parties have to engage in certain minimum costs to enter the formal process such as fixed court fees.

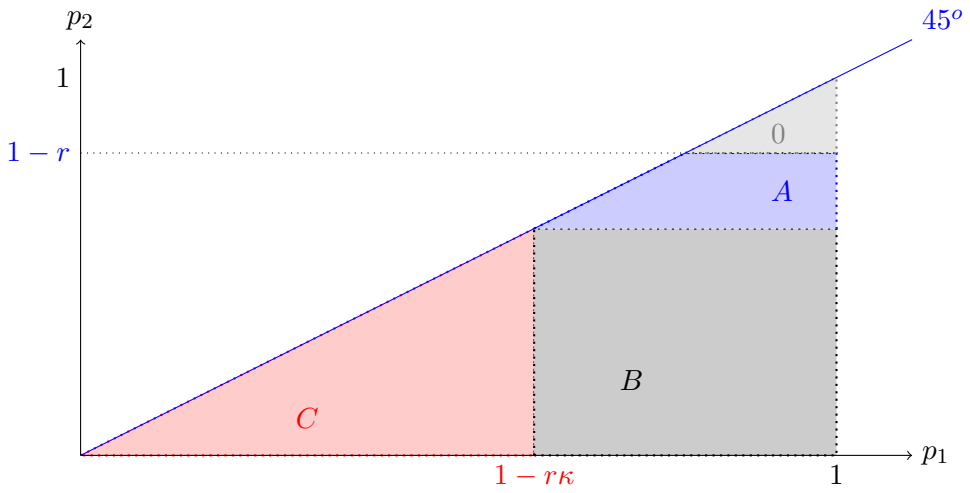

Figure 1: Partitioning the information set, given $p_1 \geq p_2$.

structures, $\mathcal{I}$, can be partitioned in the following way

$$\begin{aligned}
\mathcal{I}_0 &:= \{I \in \mathcal{I} | r > 1 - p_2\}, \\
\mathcal{I}_A &:= \{I \in \mathcal{I} | 1 - p_2 \geq r > (1 - p_2)/\kappa\}, \\
\mathcal{I}_B &:= \{I \in \mathcal{I} | (1 - p_2)/\kappa \geq r > (1 - p_1)/\kappa\}, \\
\mathcal{I}_C &:= \{I \in \mathcal{I} | (1 - p_1)/\kappa \geq r\}.
\end{aligned}$$

Figure 1 illustrates the partitioning. The equilibrium payoffs are summarized in the following lemma.

Lemma 1. *Consider the game described above and take any two probability distributions $I = (p_1, p_2)$ with $p_1 > p_2$. Then, the equilibrium payoffs $V_i(c_i)$ are*

$$V_i(1) = \begin{cases} 0 & \text{if } I \in \mathcal{I}_0 \\ 1 - r - p_2 & \text{if } I \in \mathcal{I}_A \\ (1 - p_2)\frac{\kappa - 1}{\kappa} & \text{if } I \in \mathcal{I}_B \cup \mathcal{I}_C \end{cases}$$

$$V_1(\kappa) = \begin{cases} 0 & \text{if } I \in \mathcal{I}_0 \cup \mathcal{I}_A \\ (1 - p_2 - \kappa r)\frac{\kappa - 1}{\kappa} & \text{if } I \in \mathcal{I}_B \\ (p_1 - p_2)\frac{\kappa - 1}{\kappa} & \text{if } I \in \mathcal{I}_C \end{cases}$$

$$V_2(\kappa) = 0.$$

As often in contests, the weakest type of the (ex-ante) weakest player receives zero expected payoff from participation. In addition, observe that low-cost types must obtain

the same utility since the rules of the game inevitably lead to a common upper bound on the low-cost types' equilibrium scoring support. The remaining intuition can be directly obtained via Figure 1. Region 0 corresponds to a situation in which the likelihood of meeting a high-cost player 2 is very small. Thus, a low-cost player 1 has no incentive to gamble on such an event and therefore scores high to beat a low-cost opponent. This behaviour leads to full rent dissipation. If the likelihood of meeting a high-cost player 2 is intermediate (Region A), a low-cost player 1 has an incentive to gamble on meeting a high-cost player 2. That gamble results in a reduced score which in turn leads to a reduction of the low-cost player 2's score leaving positive rents for low-cost players. If the likelihood of meeting a high-cost player 2 is relatively large (Region B and C) high-cost players enter the contest, too. Their scoring behaviour depends on the likelihood of a high-cost player 1. If this likelihood is small (Region B), a high-cost player 2 is still reluctant to invest too much as she most likely meets a low-cost player 1. This results in relatively small scores by a high-cost player 2 and thus large payoffs for a high-cost player 1. If however, the likelihood of a high-cost player 1 is similar to that of player 2, both high-cost players compete frequently reducing the payoff of a high-cost player 1. Low-cost players expected payoff remain unchanged in regions B and C as the upper bound of the bid function is only determined by player 2's type-distribution.

2.2 A Full Resolution Mechanism

One potential way for players to reduces the inefficiencies induced via the contest game is to engage in a mechanism that offers a settlement solution to parties at little or no cost. This way, players can save on the investment and increase their (expected) payoffs. The simplest such mechanism is a mechanism that offers some sharing of (the probability of winning) the prize. That is, each player receives an expected payoff of x_i independent of her type. Such a mechanism is trivially incentive compatible but may fail to meet the players participation constraint.

For the purpose of our example, assume the following grand game: Fix the above mechanism. Both players simultaneously decide whether to participate in the mechanism and if they do, the pie is split using a lottery that (in expectations) yields a payoff of x_i to each player, such that $x_1 + x_2 = 1$. If at least one player vetoes the mechanism, they play the all-pay contest. We want to sustain an equilibrium in which such a mechanism is successful with probability 1. To make a full resolution as simple as possible, we assume that the non-deviating player $-i$'s (off-path) beliefs upon observing a veto by her opponent i is a degenerate belief on $-i$ being the low-cost type. This belief is $-i$'s low-cost type's worst off-path belief, that is, it eases her participation constraint the most (for a similar argument

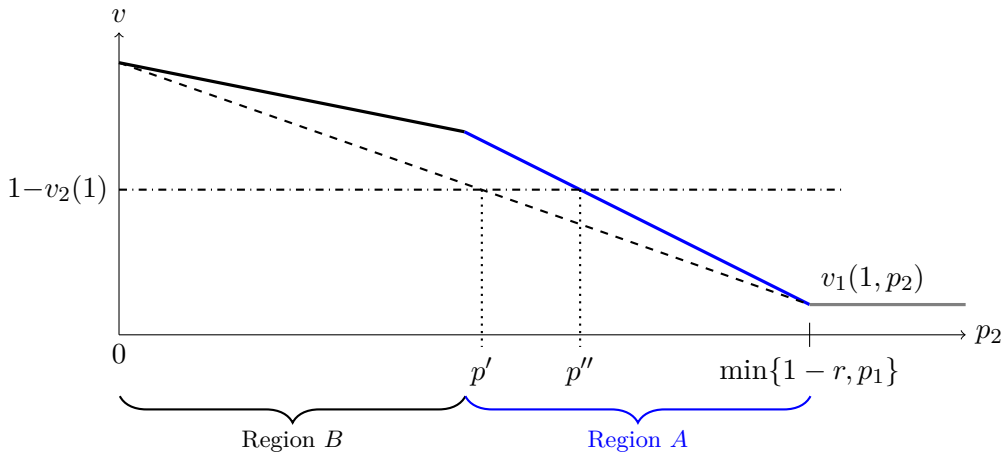

Figure 2: The value of vetoing for player 1 given player 2 assigns a probability p_1 to player 1 being the low-cost type. The dashed line denotes the functions convex hull. The dot-dashed depicts the residual resources after paying the participating share to player 2. For $\kappa = 5$, $p_1 = 1/3$ and $r = 1/6$, it follows that $p' = 7/24$ and $p'' = 1/3$.

in a more general setup see Zheng, 2017).[2]

Then, the low-cost player's value of vetoing $v_1(1)$ is her expected payoff from playing the default game under priors. The solid line in Figure 2 displays the value of vetoing as a function of the probability that player 2 has low-cost, p_2. The simple mechanism outlined above can only work if the values of vetoing add up to less than the total surplus of 1. In the graph of Figure 2 this condition fails for any $p_2 < p''$ and, in the form presented above, the mechanism is not sustainable. Numerically, the following parameter values would induce that situation.

Example. Let $\kappa = 5$, $p_2 = 7/24$, $p_1 = 1/3$ and $r = 1/6$.

Claim. Without informational punishment no full resolution mechanism exists in Section 2.2.

Proof. Let $\tilde{p}_1$ be low-cost player 1's least-preferred, possible off-path belief *of player 2 about player 1*. One such worst belief is a degenerated belief on her being the low-cost type.[3] If player 1 vetoes, both players enter the contest given the belief profile $\{\tilde{p}_1 = 1, p_2\}$, that is, they engage in region A and player 1, low cost, receives the payoff $(1 - p_2 - r) = \frac{13}{24}$. Similar, the worst off-path belief from the perspective of player 2, low-cost, $\tilde{p}_2$, is a degenerate belief on her being the low-cost type. The belief profile $\{p_1, \tilde{p}_2 = 1\}$ lies in (the mirror image of) Region A, and player 2, low-cost receives the payoff $(1 - p_1 - r) = \frac{1}{2}$. Thus, any equilibrium of a full-resolution mechanism would require a total payment of $1/2 + 13/24 > 1$ which is not feasible. □

[2] In principle we can pick any off-path beliefs for the non-deviating player. However, the deviator cannot learn anything from deviation and thus keeps the prior belief, too.

[3] Every probability mass weakly above p_2 is equally undesired.

We now augment the mechanism with *informational punishment*: The mechanism ex-ante commits to release an informative signal about the non-deviating players message in case one player vetoes the mechanism. In addition, players cannot ex-ante commit to ignore any available information. Using techniques from the Bayesian persuasion literature (Kamenica and Gentzkow, 2011), the mechanism can punish a deviator via an informative public signal. The designer uses informational punishment to relax the participation constraint and a full resolution mechanism is possible in the interval $[p', p'']$.

The most severe informational punishment is obtained by constructing a convex hull of the value of vetoing. The player's value of vetoing is then the value of this convex hull evaluated at the player's prior belief.

Claim. Consider again Section 2.2. If the designer has access to informational punishment a full-resolution mechanism exists.

Proof. The proof is constructive and conveys the intuition of informational punishment. The full resolution mechanism $\mathcal{M}^*$ is unanimously accepted, and implements the shares $(x_1, x_2) = (13/25, 12/25)$. The equilibrium of the grand game is the following: Both players accept $\mathcal{M}^*$ and report their true types. If player i vetoes, player $-i$'s (off-path) belief is $\tilde{p}_i = 1$.

In case player 1 vetoes, the mechanism releases one of two possible signal realizations, $\{h_2, l_2\}$. The probability with which a signal realizes depends on player 2's type report. The signals are constructed similar to those in Kamenica and Gentzkow (2011). A signal h_2 realizes only if player 2 consists of the high-cost type and the contest is played given the belief profile $\{\tilde{p}_1{=}1, p_2{=}0\}$. The corresponding payoff to low-cost player 1 is $4/5$. If the realization is l_2, the probability of facing a low-cost player 2 corresponds to $1 - r = 5/6$ and low-cost player 1's expected payoff is 0. Signals have to be Bayes' plausible, thus the probability of signal realization l_2 is $\rho(l_2) = 7/20$. Consequently, low-cost player 1's expected payoff is $13/25$. Similarly, if player 2 rejects she receives a signal from the set $\{h_1, l_1\}$. Signal h_2, too, is conclusive and signal l_1 results in a belief $p_1 = 1 - r = 5/6$. Bayes' rule implies that signal l_1 realizes with probability $\phi(l_1) = 6/15$ and the expected payoff of a deviating low-cost player is $12/25$. High-cost types receive a payoff of 0 when rejecting $\mathcal{M}^*$ and thus, a full-resolution mechanism exists. □

The remainder of this paper generalises the results describing how informational punishment (i) ensures full participation and (ii) lowers the participation constraint. The intuition driving the result in the example prevails under the general setup.

3 Informational Punishment

3.1 Setup

Players and Information Structure There are N players, indexed by $i \in N := \{1, ..., N\}$. Each Player is equipped with a private type $\theta_i \in \Theta_i$, with $\Theta_i \subset \mathbb{R}$ being a convex set.[4] The state $\theta := \theta_1 \times ... \theta_N \in \Theta$ is distributed according to a commonly known distribution function $F(\theta)$. Let $\theta_{-i} = \theta \setminus \theta_i$ and define the marginal $F_i(\theta_i) = \int_{\theta_{-i}} F(\theta_i, d\theta_{-i})$ and denote the support of the marginal by Q_i, i.e., $Q_i := supp(F_i(\theta)) \subseteq \Theta_i$. Note that this formulation is rich enough to cover cases of finite and convex type spaces. An information structure I is a commonly-known joint distribution over the state θ. The only restriction we impose on I is that it is absolutely continuous w.r.t. F. We refer to the information structure F as the *prior* and use the symbol I^0. Given her type, θ_i, a player's *conditional information* is $I(\theta_{-i}|\theta_i) = \frac{I(\theta_i, \theta_{-i})}{I_i(\theta_i)}$, with $I_i(\theta_i)$ the marginal of I.

Basic Outcomes and Decision Rules. There is an exogenously given set of basic outcomes, $\widetilde{Z} \subset \mathbb{R}^K$, with $K < \infty$. Player i values basic outcome $\tilde{z}$ according to a Bernoulli utility function, u_i.

Further, let π be a decision rule, that is, a mapping from the type space into a probability measure over basic outcomes. The set of attainable outcomes given the decision rule π is the product space $\widehat{Z} = \cup_I \{supp(\pi_I(dz, \theta_i, \theta_{-i}))|\forall \theta \in \Theta\} \subseteq \widetilde{Z}$, where π_I is the decision rule conditional on information structure I. Given that $\widehat{Z}$ is a compact metric space, a Borel-measurable decision rule π_I is *I-incentive compatible* if each player's von-Neumann-Morgenstern utility function , v_i, satisfies the following

$$\begin{aligned} v_i(\theta_i, I) &:= \sum_{\theta_{-i}} \int_{\widehat{Z}} u_i(\hat{z}, \theta_i, \theta_{-i}) \pi_I(d\hat{z}, \theta_i, \theta_{-i}) I^0(d\theta_{-i}|\theta_i) \\ &= \max_{\hat{\theta} \in \Theta_i} \sum_{\theta_{-i}} \int_{\widehat{Z}} u_i(\hat{z}, \theta_i, \theta_{-i}) \pi_I(d\hat{z}, \hat{\theta}_i, \theta_{-i}) I(d\theta_{-i}|\theta_i), \end{aligned}$$

almost everywhere conditional on I, that is $\forall \theta_i \in supp(I_i)$.

Default Game. In the absence of a mechanism a decision rule π^D is induced by the non-cooperative play of a given *default game*. We assume that an equilibrium for this default game exists for every I and that the set of attainable outcomes given π^D is Z, a compact metric space. Equilibrium existence implies Borel-measurability of π^D_I and the revelation principle for Bayesian games implies it is I-incentive compatible.

Mechanism. Instead of playing the default game, players can participate in a given veto-constraint mechanism offered by a non-strategic party, the designer. Independently of the

[4] Nothing of our argumentation relies on the one-dimensional type-space assumption. However, as this assumption is typically made in applications we impose it here.

mechanism's outcome space, the revelation principle implies that the equilibrium of the continuation game starting after every player accepted the mechanism induces a I^0-incentive compatible decision rule that is Borel-measurable. We therefore identify the mechanism as a mapping from the players' type reports, m, into a I^0-incentive compatible decision rule, π_{I^0}. The set of implementable decision rules might be restricted. We impose the mild assumption that $\cup_I \pi_I^D$ is included in the feasibility set.[5] The mechanism takes only place if *all* players accept it. If at least one player vetoes the mechanism by not sending a type report to it, the mechanism is void, the identities of the vetoing players become known, and the default game is played. Let I^v be some information structure players hold after observing that the mechanism has been vetoed. Then, the play of the default game implements a I^v-incentive compatible decision rule, $\pi_{I^v}^D$.

In addition to π_{I^0} the designer also proposes a signalling function Σ that maps messages of participating players into an element of a signal space $S \supseteq \Theta$. Importantly, no player has the chance to veto Σ. That is, the signalling function produces its signal independently of whether the mechanism is unanimously accepted or not.

The designer selects and commits to a (joint) mechanism $\mathcal{M} = (\pi_{I^0}, \Sigma)$ at the beginning of the game.We call $\mathcal{M}$ *a mechanism with access to informational punishment.*

Timing. The timing of the grand game is as follows: players learn their types and observe the choice of $\mathcal{M}$. Players then simultaneously decide whether to participate in the mechanism by deciding whether to send a message, $m \in \Theta_i$, to $\mathcal{M}$. If every player sends a message, then π_{I^0} is implemented. If one or more players do not send a message, the identities of these players become common knowledge and a signal $s \in S$ realises according to Σ. Thereafter players engage in the default game using an updated information structure $\bar{I}$ and payoffs realise according to $\pi_{\bar{I}}^D$.

We are looking for the set of grand mechanisms $\mathcal{M}$ implementable in a perfect Bayesian Nash equilibrium.

3.2 Informational Punishment

A general concern in veto-constraint mechanism-design problems is that the revelation principle is not valid as some allocations are only implementable if some types of some players reject the mechanism on the equilibrium path. In a first step we argue informational punishment eliminates this concern.

[5] This assumption means that a possible outcome of the designers game form is: the allocation is determined by the play of the default game (c.f., Myerson-Mediation). As a counterexample: The designer's game form has only the origin as an outcome space, i.e, when players' participate in the mechanism, they lose the possibility to play the default game, and the designer's game implements the ex-post outcome $z = 0$ with probability one.

Proposition 1. *It is without loss of generality to focus on mechanisms that ensure full participation when designing a mechanism with access to informational punishment.*

Proof. The proof is constructive. We show that for every mechanism $\mathcal{M}$ without full participation an alternative mechanism $\mathcal{M}^*$ exists, that implements the same decision rule with full-participation.

Fix any equilibrium of the grand game for some arbitrary mechanism $\mathcal{M}$ that implements a I^0-incentive compatible decision rule, π_{I^0}. Assume that at least one type of one player vetoes $\mathcal{M}$ along the equilibrium path. We refer to this equilibrium as the veto equilibrium. Let $\xi_i(\theta_i)$ denote the probability that a player type θ_i vetoes and $A{=}\{i|\xi_i(\theta_i){=}0 \ \forall \ \theta_i{\in}\Theta_i\}$ denote the set of players accepting $\mathcal{M}$ with probability 1 irrespective of their type.

We now construct a mechanism $\mathcal{M}^* = (\pi^*_{I^0}, \Sigma^*)$ providing the same expected payoff to each player as the equilibrium of $\mathcal{M}$. Thus $\pi^*_{I^0} = \pi_{I^0}$, conditional on player's accepting the mechanism.

Next, we construct Σ^* such that every player accepts $\mathcal{M}^*$. Define the random variable $S_{i,j} := \Theta_j \to \{0, 1\}$ with associated probability $Pr(1|\theta_j) = \xi(\theta_j)$. Collecting the functions $S_{i,j}$ into a vector S_i defines another random variable. Finally let Σ^* be such that whenever player i vetoes random variable S_i realises.

Finally, we show that no player is better off by unilaterally vetoing the mechanism. That is, we need to show that there is an off-path belief about a deviating player i such that i prefers to participate in the mechanism.

First assume that the mechanism $\mathcal{M}^*$ is accepted by all but player $i \in A$. Consider the following off-path continuation game: If player i rejects the mechanism, then the other players hold the same off-path beliefs about her as in they hold in when $\mathcal{M}$ is vetoed along its equilibrium path. The expected payoff at the point where player i makes her vetoing decision is the choice between the outcome $\pi^0_I(\cdot, \theta_i, \theta_{-i})$ when participating and a lottery over information structures induced by the signalling function Σ^*. However, given that she was willing to participate in $\mathcal{M}$, and Σ^* replicates the post-veto distribution given the equilibrium in $\mathcal{M}$, choosing the lottery can – by construction – not be preferred to the expected outcome implemented by $\pi^*_{I^0}$. Thus, player i has an incentive to participate.

Next consider the situation of a player $i \notin A$. If such a player vetoes under $\mathcal{M}$, she faces opponents with an equilibrium belief according to the veto information structure $\bar{I}^{\mathcal{M}}_i$. Under $\mathcal{M}^*$ the belief on a vetoing $i \notin A$ is off-the equilibrium path and we can specify that belief to be the same as $\bar{I}^{\mathcal{M}}_i$ which is Bayes' plausible by construction. □

The intuition of this result can be directly observed using the example from Section 2. On-path rejection of the mechanism is only beneficial to the designer if it relaxes the implementability constraints, in particular the participation constraints. The participation con-

straints are a particular concern if the outside option to the mechanism is a non-cooperative Bayesian game. In such case the value of that outside option is endogenous as it depends on the information structure after a veto. On-path vetoing relaxes the constraints in so far as that it splits the prior information structure into two parts: (i) the information structure after veto; and (ii) that after acceptance. Both information structures have to be Bayes' plausible, that is, the prior is a *combination* of the two in the sense of Bergemann and Morris (2016a). If the information structure for an on-path veto can now prohibit other, *accepting players* from vetoing (as it reduces their value of vetoing), on-path rejection may be beneficial. A designer that has access to informational punishment can, however, induce such an information structure without requiring on-path rejection. Instead, the designer can promise to carry out an experiment on the complying player and to report the outcome of this experiment publicly.

Under Proposition 1 the designer can freely pick the off-path beliefs when designing the full participation mechanism, that is, he also acts as equilibrium selection device. In some applications such an option may not be available. In fact it might actually be the case that a refinement criterion that limits the designers choice of the equilibrium *causes* the necessity of on-path vetoes.

Consider now, that the designer's choice of equilibrium selection, that is, off-path beliefs, is restricted by a refinement criterion

$$(\star) \in \{ \text{ Perfect Sequential Equilibrium, Intuitive Criterion, Ratifiability}\}.$$

Our second statement shows that such a restriction on the off-path belief does not change the result of Proposition 1.

Proposition 2. *Suppose the solution concept is perfect Bayesian equilibrium with refinement concept* $(\star)$*. It is without loss of generality to assume full participation when designing a mechanism with access to informational punishment.*

Proof. Suppose the hypothetical veto equilibrium in Proposition 1 satisfies refinement $(\star)$. We want to show that the full-participation equilibrium with $\mathcal{M}^*$ satisfies the same refinement criterion. Take any $i \in =\{i|\xi_i(\theta_i)=0 \ \forall \ \theta_i \in \Theta_i\}$ and an arbitrary θ_i. Consider the continuation game that begins with θ_i's decision whether to accept, action 1, or to veto, action 0, $\mathcal{M}^*$. By construction, both actions lead to the same (expected) continuation equilibrium payoffs as in the hypothetical vetoing equilibrium. Thus, her choice is isomorphic to the situation in the veto-equilibrium and any credible (according to $(\star)$) inference her opponent can draw from her decision is the same in both games. As the decision rule, π^E, that is implemented in the veto equilibrium is ratifiable (respectively: the veto equilibrium is perfect sequential or satisfies the intuitive criterion), the very same decision rule,

π^E, that is implemented in the full-participation equilibrium is ratifiable (respectively: the full-participation-equilibrium is perfect sequential or satisfies the intuitive criterion).

Now, take $i \notin = \{i | \xi_i(\theta_i) = 0 \ \forall \ \theta_i \in \Theta_i\}$ and fix an arbitrary θ_i. Perfect sequential rationality and ratifiability of the implemented decision rule π^E, hold by construction for any such player with the off path belief used in the proof of Proposition 1. That belief is a credible veto belief and makes every type in the support indifferent between accepting and vetoing.Finally, we show that the same holds for the intuitive criterion. Let the set of types that weakly prefer vetoing $\mathcal{M}^*$ for *some* belief be T . Full-participation fails the Intuitive Criterion if the set of types that *strictly* prefer rejection for *all* belief with support on T is non-empty. Call this set $\widehat{T}$. We now argue that $\widehat{T} = \emptyset$. By construction, every type in the support of the off-path belief in Proposition 1, $\tilde{p}_i$, is in T. Thus, $supp(\tilde{p}_i) \subset T$. However, for $\tilde{p}_i$, every type in $supp(\tilde{p}_i)$ does not *strictly* prefer rejection. Thus $\widehat{T} \setminus supp(\tilde{p}_i) = \Theta^2$. Moreover, consider $\theta_i \in \Theta_i \setminus supp(\tilde{p}_i)$. Even if $\theta_i \in T$, that is this type profits for some beliefs, θ_i cannot be in $\widehat{T}$, as θ_i weakly prefers to accept the mechanism: otherwise this type would have vetoed the mechanism in the veto-equilibrium. Because $supp(\tilde{p}_i) \vee \Theta_i \setminus supp(\tilde{p}_i) = \Theta_i$ it must be the case that $\widehat{T} = \emptyset$. □

However, informational punishment may be beneficial even if full participation is optimal. The Bayesian persuasion experiments we construct for potential deviation operate directly on the players participation constraint. That is, even if full participation is guaranteed at the optimum, then informational punishment may still be useful to the designer if the following conditions are satisfied (a) the participation constraint at the optimum is binding, and (b) the value of vetoing is concave around the prior. In such a case the mere promise of a persuasion experiment reduces that player's *effective* value of vetoing. Let $\mathcal{I}_{F_i}$ be the set of all information structures in which the distribution over types of player i is fixed to F_i, and let $I_{F_i} \in \mathcal{I}_{F_i}$ denote a generic element in that set. Then, given any off-path belief F_i^v on a deviator i, informational punishment can reduce the value of vetoing to the largest function weakly smaller than the original v_i that is convex in $I_{F_i^v}$.

Corollary 1. *The optimal mechanism with informational punishment is outcome equivalent to the optimal mechanism without informational punishment and a convexified value of vetoing,*

$$\hat{v}_i(\theta) := \max\{f : I_{F_i^v} \to \mathbb{R} | f \, convex \, and \, f(I_{F_i^v}) \leq v_i(\theta, I_{F_i^v})\}.$$

The result follows directly from Proposition 1, but allows for an alternative interpretation: Given any function v_i and a mechanism design problem with access to informational punishment, we can reduce each players participation constraint at no loss to $\hat{v}(\theta, I^0_{F_i^v})$ if we are only interested in the outcome of the optimal mechanism. Figure 3 provides a graphical intuition for the result.

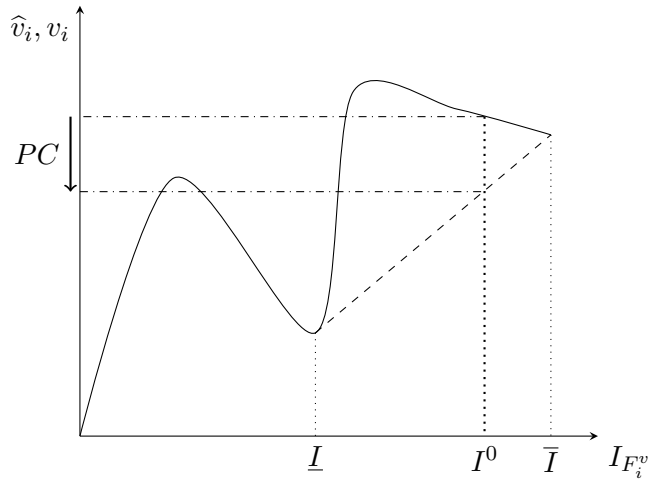

Figure 3: Value of vetoing and convex hull. *Given the prior information structure player i's participation constraint can be reduced by promising to realise a signal providing a mean-preserving spread over the prior and either changing the belief to either $\underline{I}$ or $\overline{I}$.*

3.3 Informed Principle Problems

Next, consider the following extension of our setting. Instead of a non-strategic designer, one of the players, say $i = 0$, proposes a mechanism. The setting becomes an informed-principal problem. A key concept to solve informed-principal problems is the concept of inscrutability (see Myerson, 1983).[6] As pointed out by Celik and Peters (2011) the principle of inscrutability might fail in a setting as the one considered in this paper. Our final result states that if we allow player 0 to propose mechanisms that are augmented by signals, then the principle of inscrutability is satisfied.

Proposition 3. *In an informed principal setting, if the principal can propose mechanisms that allow for informational punishment, then the principle of inscrutability is satisfied.*

Proof. Suppose there exists an equilibrium such that different types of player 0 propose different mechanisms. Let M be the of mechanisms that are proposed with positive probability. Let $\xi_0^m(\theta_0)$ denote the probability that principal type θ_0 proposes mechanism $m \in M$.

Fix any equilibrium of the grand game described above. The equilibrium play of the grand game implements a I^0 incentive-compatible decision rule $\pi_{I^0}^E$. Assume that at least one type of one player vetoes at least some $m \in M$ along the equilibrium path. We refer to this equilibrium as the separating-and-veto equilibrium. For $i > 0$, let $\xi_i(\theta_i)$ denote the probability that a player type θ_i vetoes and $A{=}\{i|\xi_i(\theta_i){=}0 \ \forall \ \theta_i{\in}\Theta_i\}$ denote the set of players accepting $\mathcal{M}$ with probability 1 irrespective of their type.

[6]The principle of inscrutability is essential for virtual all solution approaches to informed-principal problems (see for example Maskin and Tirole, 1990, 1992; Mylovanov and Tröger, 2014).

We now construct a mechanism $\mathcal{M}^* = (\Pi^*, \Sigma^*)$ providing the same expected payoff to each player as the proposed equilibrium. Thus $\Pi^* = \pi^E_{I^0}$, conditional on players accepting the mechanism.

Next, we construct Σ^* such that every player accepts $\mathcal{M}^*$. For $j = 0$ and $i > 0$, define the random variable $S_{i,0} := \Theta_0 \to M$ with associated probability $Pr(m|\theta_0) = \xi^m_0(\theta_0)$. For any $i, j > 0$, define the random variable $S_{i,j} := \Theta_j \to \{0, 1\}$ with associated probability $Pr(1|\theta_j) = \xi(\theta_i)$. Collecting the functions $S_{i,j}$ into a list S_i defines another random variable. Finally let Σ^* be such that whenever player i vetoes random variable S_i realises.

Finally, we show that no player is better off by unilaterally vetoing the mechanism. That is, we need to show that for every player $i > 0$ there exists an off-path belief the other players hold about i such that i prefers to participate in the mechanism.

First assume that the mechanism $\mathcal{M}^*$ is accepted by all but player $i \in A$. Consider the following off-path mechanism: If player i rejects the mechanism and $S_{i,0}$ realizes, then the other players hold the same off-path beliefs about her as in the hypothetical separation-and-veto equilibrium after player 0 proposed the mechanism $m = S_{i,0}$. The expected payoff at the point where player i makes her vetoing decision is the choice between the outcome $\pi^E(I^0, \theta_i, \theta_{-i})$ when participating and a lottery over information structures induced by the signalling function Σ^*. However, given that she was willing to participate in every $m \in M$, and Σ^* replicates the post-veto distribution given the equilibrium under E, choosing the lottery can – by construction – not be preferred to $\pi^E_{I^0}$. Thus, player i has an incentive to participate.

Next consider the situation of a player $i \notin A$. The situation is similar for such a player, with the differences that a veto implies off-path beliefs on that player which are on-path beliefs in the veto equilibrium.

□

4 Discussion

In this section we discuss our crucial assumptions, the implications on the results and several applications.

The most crucial assumption we make is that players can on the one hand publicly veto a mechanism, but are unable to commit to ignoring information. An important consequence of public veto ability is that the mechanisms power in the event of vetoing is limited. The simplest way to see how this limits the mechanisms power is to consider a basic coordination mechanism a la Myerson (1982). In equilibrium of such a mechanism, the designer sends recommendations about actions to the players who subsequently play a pre-defined game obeying these recommendation. Any off-path behaviour such as sending an empty message

can be punished by the mechanism by committing to sending the *worst on-path recommendations* for the deviating player. As the opponents do not know about the deviation, she is going to obey the recommendation and thus punishing the deviator. Such a behaviour is *not* possible in a veto-mechanism. Here, the deviator can credibly proof that she did not send an on-path message and therefore trigger an off-path game both players are aware of entering.

An equivalent assumption to verifiable non-participation is that players can at an *interim stage* commit to not participate for example by allowing public monitoring of their (non-existing) communication with the mediator. At the same time players cannot commit at any point in time on ignoring publicly available information. That is, although the player would like to announce that she is not going to participate in the mechanism and is going to ignore any information provided by the mechanism in the future, players are going to assume that the announcement of ignorance is pure cheap talk. This way, even a non-participating player has an incentive to listen to the information once it is publicly available.

We argue that the commitment structure that we assume indeed resembles real-world scenarios. The main reason is that it remains optimal and possible to make a veto public even *after* it has been made. For example, in vertical contracts firms can make a public statement of non-negotiation that is legally binding or legal disputants can verify absence at a council meeting. At the same time we consider it unlikely for a player to be able to commit to ignoring information that is available and beneficial at an interim stage.

A second assumption we make is that the mechanism in fact has the legal power to disclose information despite not being able to act otherwise. Thus, a necessity for informational punishment is that mechanisms are exempt from non-disclosure laws at least in off-path events. Note that this assumption is substantially less demanding than that made in "trembling mechanisms" (Gerardi and Myerson, 2007; Correia-da-Silva, 2017). A trembling mechanism "fails" with positive probability although all parties vow to cooperate. Informational punishment, to the contrary, exclusively operates on an off-path event thereby not causing any violations of privacy with positive probability. Thus, all we require is the ability to commit to a signalling function in case the game enters an off-path note.

We neither impose assumptions on the underlying default game nor on the outcome space of the mechanism. Thus, when facing a mechanism design problem it is without loss of generality to convexify the participation constraint and assume full participation, which in turn allows straight-forward application of well-known simplifications via the revelation principle or the principle of inscrutability. Furthermore, our approach is constructive and characterizes the structure of informational punishment that relaxes the participation constraint as much as possible.

5 Conclusion

We provide a simple extension to the classic mechanism design framework to characterize the optimal mechanism in the presence of a Bayesian outside option to the mechanism. We show that if players cannot commit to ignoring public information, the optimal mechanism can relax player's participation constraint by threatening them to release a public signal about the information obtained from participating players. This threat of informational punishment is sufficient to restrict attention to full-participation mechanisms. Informational punishment reduces the players value of vetoing to its convex hull with respect to the information structure, but does not effect incentive compatibility constraints or the designer's on-path expected payoff. Using informational punishment therefore allows to use standard methods of mechanism design to characterize the optimal mechanism in the presence of an arbitrary Bayesian default game.

Using informational punishment we can restore classic general results such as the revelation principle with full-participation and the principle of inscrutability. Our findings allow tractable solutions for a variety of applied problems. Veto mechanism and Bayesian outside options are present in many areas of law and economics, industrial organization, and finance. Often institutions in these areas also have the option to provide all relevant players with a public signal which is the only necessary condition for informational punishment. Thus, institutional design in these areas is equipped with informational punishment. Using our findings, abstraction from Bayesian interaction in case of failure to coordinate does not come at a large loss of tractability anymore allowing a straightforward application of such an outside option and thus a simple, and complete characterization of the incentives at work.

References

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

Balzer, Benjamin (2016). "Collusion in Auctions: An informed principle perspective". *mimeo*.

Balzer, Benjamin and Johannes Schneider (2017a). "Managing a Conflict". *mimeo*.

— (2017b). "Optimal Alternative Dispute Resolution". *mimeo*.

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

Bergemann, Dirk and Stephen Morris (2016b). "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium". *American Economic Review* 106, pp. 586–91.

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

Bolton, Patrick and David A Skeel (2010). "How to Rethink Sovereign Bankruptcy A New Role for the IMF?" *Overcoming developing country debt crises*. Ed. by Barry Herman, José Antonio Ocampo, and Shari Spiegel. OUP Oxford, pp. 449–486.

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

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

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

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

Hart, Oliver and Jean Tirole (1990). "Vertical integration and market foreclosure". *Brookings papers on economic activity. Microeconomics* 1990, pp. 205–286.

Hörner, Johannes, Massimo Morelli, and Francesco Squintani (2015). "Mediation and Peace". *The Review of Economic Studies* 82, pp. 1483–1501.

Jehiel, Philippe and Benny Moldovanu (2001). "Efficient design with interdependent valuations". *Econometrica* 69, pp. 1237–1259.

Kamenica, Emir and Matthew Gentzkow (2011). "Bayesian Persuasion". *American Economic Review* 101, pp. 2590–2615.

Lerner, Josh and Jean Tirole (2004). "Efficient patent pools". *The American Economic Review* 94, pp. 691–711.

— (2015). "Standard-essential patents". *Journal of Political Economy* 123, pp. 547–586.

Maskin, Eric and Jean Tirole (1990). "The principal-agent relationship with an informed principal: The case of private values". *Econometrica: Journal of the Econometric Society*, pp. 379–409.

— (1992). "The principal-agent relationship with an informed principal, II: Common values". *Econometrica: Journal of the Econometric Society*, pp. 1–42.

McAfee, R Preston and John McMillan (1992). "Bidding rings". *The American Economic Review*, pp. 579–599.

McAfee, R Preston and Marius Schwartz (1994). "Opportunism in multilateral vertical contracting: Nondiscrimination, exclusivity, and uniformity". *The American Economic Review*, pp. 210–230.

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

— (1983). "Mechanism design by an informed principal". *Econometrica: Journal of the Econometric Society*, pp. 1767–1797.

Mylovanov, Tymofiy and Thomas Tröger (2014). "Mechanism design by an informed principal: Private values with transferable utility". *The Review of Economic Studies* 81, pp. 1668–1707.

Philippon, Thomas and Vasiliki Skreta (2012). "Optimal Interventions in Markets with Adverse Selection". *American Economic Review* 102, pp. 1–28.

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

Zheng, Charles (2017). "A necessary and sufficient condition for Peace". *mimeo*.