Mechanism Design with Informational Punishment*

Benjamin Balzer† Johannes Schneider‡

November 2, 2020

Abstract

We introduce *informational punishment* to the design of mechanisms that compete with an exogenous status quo mechanism: The designer can publicly communicate with all players even if some decided not to communicate with the designer. Optimal informational punishment ensures that full participation is optimal and thereby restores the revelation principle.

*Johannes Schneider gratefully acknowledges support from the Ministerio Economía y Competitividad, grant ECO2017-87769-P, and Comunidad de Madrid, MAD-ECON-POL-CM (H2019/HUM-5891)
†University of Technology Sydney, benjamin.balzer@uts.edu.au
‡Universidad Carlos 3 de Madrid, jschneid@eco.uc3m.es

1 Introduction

A common theme in mechanism design is to look for a *mechanism* that outperforms the *status quo*, e.g. the current institution. Often, however, this change requires the consent of the parties involved. If the status quo is a game itself, agents' outside options are endogenous and there may not be an optimal mechanism in which all types of all players participate (see e.g. Celik and Peters, 2011). Different from mechanisms in which players' outside option are an exogenously given type-dependent value, the classical revelation arguments a la Myerson (1982) do not hold. Assuming full participation is not without loss anymore. The reason is that refusal to participate carries a signaling value. To mitigate the issue the literature's focus was on stochastic mechanisms that 'self-reject' (see Gerardi and Myerson, 2007; Correia-da-Silva, 2020): the designer enforces the status quo randomly to reduce the value of a rejection.

In Balzer and Schneider (2020) we provide an alternative approach. We show in an example how the threat of providing information about compliers can improve the acceptance of take-it-or-leave-it offers when rejections can be made publicly.[1] In this note, we show that this idea of *informational punishment* applies to design problems with endogenous outside options in general. If the designer can publicly communicate information obtained from the parties even if the mechanism is rejected, then it is without loss to assume full participation at the optimum. The revelation principle holds.

2 Setup

Players and Information Structure. There are N players, indexed by $i \in \mathcal{N} := \{1, ..., N\}$. Each player has a private type $\theta_i \in \Theta_i$ and $\Theta_i \subset \mathbb{R}$ is compact. The state $\theta := \theta_1 \times ... \times \theta_N \in \times_i \Theta_i =: \Theta$ is distributed according to a commonly-known distribution function $I^0(\theta)$, the *prior information structure*. Let $\theta_{-i} := \theta \setminus \theta_i$, and define the marginal $I^0_i(\theta_i) := \int_{\Theta_{-i}} I(\theta_i, d\theta_{-i})$ with support $supp(I^0_i) \subseteq \Theta \setminus \Theta_i$.

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. I^0, that is, $supp(I) \subseteq supp(I^0)$. Given θ_i a player's *belief* about the other players' types is the conditional distribution $I(\theta_{-i}|\theta_i) = \frac{I(\theta_i, \theta_{-i})}{I_i(\theta_i)}$, where $I_i(\theta_i)$ is the marginal of I. Let $\mathcal{I}^0$ be the set of all information structures for which I^0 is an expansion. That is, $I \in \mathcal{I}^0$ if and only if there exists a random variable $\widetilde{\Sigma}$ which maps types into distributions of signals such that the realization σ together with $\widetilde{\Sigma}$ and I^0 implies I via Bayes' rule.

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

[1] Self-rejection mechanisms have no bite in that case.

We represent the rules of a game by a decision rule, $\pi : \Theta \to \Delta(Z)$. This rule is a mapping from *reports* to probability measures over outcomes, $\mu_\pi(\cdot|\theta) = \pi(\theta)$.

Status quo. The status quo is an exogenous game of incomplete information. We assume an equilibrium in that game exists for any information structure $I \in \mathcal{I}^0$ and take the equilibrium selection as given. For any information structure I, the status quo induces a decision rule π_I^M. Under π_I^M the expected utility of a truthfully reporting player i, type θ_i is

$$\begin{aligned} v_i(\theta_i, I, \pi_I^M) &:= \int_{\Theta_{-i}} \int_Z u_i(z, \theta_i, \theta_{-i}) d\mu_{\pi_I^M}(z|\theta_i, \theta_{-i}) I^0(d\theta_{-i}|\theta_i) \\ &= \max_{m_i \in \Theta_i} \int_{\Theta_{-i}} \int_Z u_i(z, \theta_i, \theta_{-i}) d\mu_{\pi_I^M}(z|m_i, \theta_{-i}) I(d\theta_{-i}|\theta_i), \end{aligned}$$

almost everywhere conditional on I, that is, $\forall \theta_i \in supp(I_i)$. The second line follows because π_I^M is incentive compatible under information structure I (I-IC henceforth). Truthful reporting is optimal for all types of all players given π_I^M.

Existence of equilibrium under $\mathcal{I}^0$ implies that the collection of possible status quo outcomes, $\Pi^M := \{\pi_I^M\}_{I \in \mathcal{I}^0}$ with π_I^M being I-IC, is well-defined.

Mechanism. The mechanism is an alternative to the status quo. Any mechanism is a game of incomplete information represented by a decision rule. The collection of decision rules is Π. Given π and I, we define each player's optimal reporting strategy $m_{i,I}(\theta_i)$. We collect players' reports in $m_I(\theta)$. An equilibrium of π implements the decision rule $\pi_I := \pi \circ m_I : \Theta \to \Delta(Z)$ which is I-IC. [2]

The set of available mechanism, Π, may be restricted by legal or institutional constraints, or certain outcomes may simply be infeasible. We assume two minimal requirements on the set of available mechanism:

(i) $\Pi^M \subseteq \Pi$, and

(ii) Π is closed under convex combinations, that is, if $\pi, \pi' \in \Pi$, then for any $\lambda : \Theta \to [0,1]$ $\lambda\pi + (1-\lambda)\pi' =: \pi^\lambda \in \Pi$.

The first property implies that the mechanism can replicate the status quo. The second property implies that if two games (1 and 2) are part of the available mechanism, so is the game in which game 1 is played for certain type reports and game 2 for the remaining type reports.

Apart from these requirements we do not restrict Π. We allow for both a classical mechanism-design setting and the possibility that the designer's set of mechanisms is exogenously restricted. The latter includes pure 'mediation' within the status quo.

[2] Although any decision rule in Π represents a direct-revelation mechanisms, truthful implementation may not be guaranteed. Indeed, Π is shorthand for all game forms with m_i a player's action. The equilibrium play of each $\pi \in \Pi$ under I then induces some I-IC decision rule.

Informational Punishment. We assume all players have access to a signaling device Σ. The N-dimensional random variable $\Sigma : \Theta \to S$ maps type reports into realizations in signal space $S \equiv S_1 \times S_2 \times .. \times S_N$ with $|S_i| \geq |\Theta_i|$. We denote the realization of Σ by $\sigma \in S$, that of element Σ_i by $\sigma_i \in S_i$.

Timing. First, players learn their types and observe π. Second, they simultaneously send a message m_i^{Σ} to Σ. Third, players simultaneously decide whether to veto the mechanism. If at least one player vetoes the mechanism, the set V of vetoing players becomes common knowledge and the signal realizations σ become public. Players use that information to update to an information structure $I^{V,\sigma} \in \mathcal{I}^0$ and $\pi^M_{I^{V,\sigma}}$ is implemented via the status quo. If players unanimously ratify the mechanism, they report m_i to the mechanism which implements π.

Solution Concept and Veto Beliefs. We consider all mechanisms implementable as a perfect Bayesian equilibrium (PBE) of the grand game.

We make use of *veto information structures*, I^V: the information structure that arises *after* an observed veto, but *before* the realization σ. PBE implies that $I^V(\theta_{-i}|\theta_i) = I^{V\setminus i}(\theta_{-i}|\theta_i)$ for any $i \in V$ and $I^V(\theta_{-i}|\theta_i) = I^{V \cup i}(\theta_{-i}|\theta_i)$ for any $i \notin V$. In addition, all but first-node of-path beliefs on deviators are derived via Bayes' rule. The remaining off-path beliefs are arbitrary.[3]

3 Analysis

Absent informational punishment an optimal full-participation mechanism may not exist, even if Π is large. Consider the case in which players are expected to participate. A deviator i who vetoes guarantees herself the prior belief about the other players through that deviation. At the same time she can at most be punished by the "worst" off-path belief assigned to her. If i's outside option exhibits concavities in the information structure, full participation *without* informational punishment is not always optimal. The designer can further relax participation constraints by providing incentives for on-path rejections (see Celik and Peters (2011) for details).

With informational punishment the designer can relax i's participation constraint without the need of on-path rejections and (possibly) beyond what is implementable with rejections.

Proposition 1. *It is without loss of generality to focus on mechanisms that ensure full participation if informational punishment is available.*

Proof. The proof is constructive. Take any π, and a *veto equilibrium* in which the mechanism is vetoed with positive probability on the equilibrium path. We first characterize

[3] In our setting a player is at most observed to deviate once. Off-path belief cascades are thus not possible in our model.

the decision rule of the veto equilibrium. Then, we show it can be implemented with full participation using *informational punishment*.

Let $\xi(\theta)$ be the probability that π is vetoed given type profile θ. Moreover, $\xi_i(\theta_i)$ is the likelihood that type θ_i vetoes π on the equilibrium path. The set of players that vetoed, V, might be random. After a veto, players observe the set of players V_k that vetoed and update to information structure I^{V_k}. Outcomes realize according to $\pi^M_{I^{V_k}}$. Taking expectations over all realizations of V, V_k, the ex-ante expected continuation game conditional on a veto is a lottery $(P(V_k), \pi^M_{I^{V_k}})$ defined over all V_k. $P(V_k)$ is the on-path likelihood that a veto is caused by the set V_k and not by any other set. Because $\Pi^M \in \Pi$ and Π is closed under convex combinations, the lottery implies $\pi^M_{I^{\mathbb{E}V_k}} = \sum P(V_k)\pi^M_{I^{V_k}} \in \Pi$.

Conditional on no veto, the information structure is I^a, and π_{I^a} is the decision rule.

The grand game implements an I^0-IC decision rule $\pi'_{I^0} := \xi\pi^M_{I^{\mathbb{E}V_k}} + (1-\xi)\pi_{I^a}$. Again, $\pi' \in \Pi$ because Π is closed under convex combinations.

We now construct a signal Σ such that the mechanism π'_{I^0} is implementable under full participation. By construction, π' is feasible and I^0-IC. What remains is to show that no player has an incentive to veto π'.

We construct the following signaling device $\Sigma_i : \Theta_i \to \Delta(\{0,1\})$ where $\sigma_i(\theta_i) = 1$ with probability $\xi_i(\theta_i)$ and 0 otherwise. When observing off-path behavior (i.e., a veto) by i, j believes that i uniformly randomized over the entire type-space when reporting to Σ_i. Thus, she disregards σ_i. We choose the off-path belief on i identical to the belief attached to i after observing her unilateral veto in the veto equilibrium.

No player i has an incentive to veto the mechanism. If a player vetoes the mechanism Σ_i provides her with the same lottery over information structures that she expects from a veto in the veto equilibrium. Participation, in turn, gives the same outcome as the veto equilibrium. No player can improve upon the outcome of the veto equilibrium by vetoing π'.

Truthful reporting to Σ_i is a best response. Σ_i is on-path payoff-irrelevant. Thus, under (π', Σ) an equilibrium with full participation in π' exists that implements the same outcome as the veto equilibrium. □

4 Conclusion

We show that the threat to partially reveal information about complying players can make deviations less attractive. As a result—given minimal conditions on the available mechanism—the revelation principle holds and full participation is optimal. Informational punishment works off the equilibrium path, does not affect the mechanism directly, and allows for publicly verifiable rejections. It is straight-forward to show that informational punishment can be decentralized and that the result of Proposition 1 extends to standard belief refinements and ensures inscrutability in informed principal problems.[4]

[4]Due to space constraints we do not present these results here, they are available upon request.

References

Balzer, B. and J. Schneider (2020). “Persuading to Participate: Coordinating on a Standard”. *mimeo*.

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

Correia-da-Silva, J. (2020). “Self-rejecting mechanisms”. *Games and Economic Behaviour*, pp. 434–457.

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

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