Soc Choice Welfare (2006) 26: 455–470
DOI 10.1007/s00355-006-0091-5

ORIGINAL PAPER

David P. Baron · Adam Meirowitz

Fully-revealing equilibria of multiple-sender signaling and screening models

Received: 21 January 2003 / Accepted: 28 May 2004 / Published online: 12 April 2006

Abstract Models of incomplete information have played a major role in the fields of political science and political economy. The models have almost exclusively been signaling models, and their substantive focus is frequently on situations in which the extensive form is not dictated by institutional requirements or procedures. We explore the relation between multiple-sender signaling games and the corresponding screening or mechanism design games without transfers and establish an equivalence result. If there is a fully-revealing equilibrium in the signaling game there is also a full-information optimal mechanism that yields the principal's optimal policy in every state. The converse, that fully-revealing equilibria exist in the signaling game if a full-information optimal mechanism exists, is true if and only if the mechanism involves only the selection of policies that are optimal for some belief about the state. We also present two straightforward sufficient conditions for the existence of full-information optimal mechanisms. When either holds, fully-revealing equilibria in the signaling and screening games exist. The perceived advantage of the signaling over the screening approach – that no commitment by the principal is assumed – may be over-stated as flexibility in specifying off the path beliefs can mimic commitment.

1 Introduction

In an environment with informed and uninformed players, two basic approaches can be taken to characterize their strategic interactions. In the signaling approach, the

D. P. Baron (✉)
Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA
E-mail: baron_david@gsb.stanford.edu

A. Meirowitz
Department of Politics, Princeton University, Princeton, NJ, USA
E-mail: ameirowi@princeton.edu

informed players, the agents or senders, move first and the uninformed player, the principal or receiver, then chooses a policy. In the screening or mechanism design approach, the uninformed player moves first and the informed players move next. In the mechanism design approach, the principal is typically assumed to commit to a policy for any action the informed players might take. This commitment is to be understood as arising from third-party enforcement or from some repeated play equilibrium that supports the mapping from the agents' actions to a policy. Neither repeated play nor third-party enforcement is typically incorporated in the game, so, commitment is assumed. The choice between modelling a strategic situation using the signaling versus screening approach is consequential because the equilibria can be different.

In the fields of political science and political economy the substantive focus of these models is frequently on situations in which the extensive form is not dictated by institutional requirements or procedures. Lobbying, for example, is a relatively unstructured process. The study of legislative organization is also unstructured as the US Constitution does not specify how the chambers of Congress are to be organized. In these fields the analysis has almost exclusively relied on signaling models. Signaling models have been preferred for two reasons. First, in contrast to markets, in political settings utility is typically non-transferable, and screening models generally assume transferable utility. Second, in political settings the commitment assumption of the screening approach has been viewed as strong. Moreover, recent results of Battaglini (2002) and Krishna and Morgan (2001) show that fully-revealing equilibria exist in certain signaling games.

In this paper we consider the relation between fully-revealing equilibria of multiple-agent signaling and screening games without transfers.[1] We show that in some situations, the commitment assumption of the screening approach is mimicked by flexibility in selecting off the equilibrium path beliefs in the signaling approach. One implication of this finding is that the perceived advantage of the signaling approach – avoiding the assumption of commitment–may be overstated in games with fully-revealing equilibria. The paper presents conditions on the informational environment that are sufficient for fully-revealing equilibria in both the signaling and screening representations.

The equivalence results because even though screening formulations incorporate explicit commitment that signaling games do not assume, the perfect Bayesian equilibrium concept often used for signaling games with infinite message and action spaces can replicate commitment because it allows a degree of flexibility in specifying off-the-equilibrium-path beliefs. That is, beliefs are unrestricted off the equilibrium path, provided there is an action that is sequentially rational given those beliefs.[2] This allows the modeler considerable flexibility in choosing off-the-equilibrium-path beliefs, which may make it possible to support a fully-revealing equilibrium. Choosing the beliefs in response to off-the-equilibrium-path

[1] McAfee and Reny (1992) and Cremer and McLean (1985, 1988), however, show that if the information of the agents is correlated the principal can achieve her full-information optimum. Their theory, and the mechanism design approach more generally, typically assumes that utility is transferable between the principal and the agents. When utility is non-transferable, the full-information optimum may not be attainable (Melumad and Shibano 1991).

[2] The refinements used to select equilibria in finite signaling games are usually not applied to signaling games with infinite message and action sets, so beliefs off the equilibrium path are unrestricted.

messages then can be equivalent to committing to sequentially rational actions as a function of the messages, as in mechanism design. This suggests a relation between full-information optimal mechanisms and fully-revealing signaling equilibria.

Gilligan and Krehbiel (1987, 1989) analyzed "open rule" models of legislatures in which the state and policy spaces are unidimensional and agents have quadratic preferences over the outcome, which depends on a policy choice and a random shock. As in Crawford and Sobel (1982) they showed that with an open rule there are no fully-revealing equilibria in a signaling model (a cheap talk game) with one sender.[3] In their 1989 paper, they considered the same model with two equally-informed senders, majority and minority members of a committee, and characterized a partially-revealing equilibria. Krishna and Morgan (2001) reconsidered the two-sender, open-rule model and identified a fully-revealing equilibrium.[4] As Krehbiel (2001) and Battaglini (2002) pointed out, the fully-revealing equilibrium identified by Krishna and Morgan (2001) involves off-the-equilibrium-path beliefs that are sharply discontinuous in the messages of the informed agents.[5] These beliefs then have best-responses for the receiver that punish the agents in a manner sufficient to induce them to fully reveal their information. The partially-revealing equilbirium characterized by Gilligan and Krehbiel (1989) also involves this type of discontinuity.

Battaglini (2002) considered the case in which the policy and state space have two or more dimensions and established a robust fully-revealing equilibrium to the signaling game with equally-informed agents. His construction relies on the local properties of the agents' indifference curves at the principal's ideal point. When equally-informed agents have strictly quasi-concave utility functions, the implicit alignment of principal and agent preferences over subspaces of the outcome space is used to generate fully-revealing equilibria. While robust and coalition proof, Battaglini's equilibrium in the two-dimensional model involves beliefs on the part of the principal that ignore some of the information that each agent reports. Specifically, the principal relies on each agent's message (which is a two-dimensional vector) for only one coordinate, and since the principal has no incentive to deviate, it is a perfect Bayesian equilibrium. The principal, in effect commits to ignore part of the message of each agent. The results developed here provide insight into these fully-revealing signaling equilibria and their relation to equilibria of the corresponding screening game.

The paper is organized as follows. Section 2 identifies the class of environments considered. Section 3 presents results on the relation between the existence of full-information optimal mechanisms and fully-revealing perfect Bayesian equilibria. Section 4 presents sufficient conditions for the existence of these types of equilibria within the class of environments considered. Section 5 concludes with a discussion.

---

[3] See also Austen-Smith and Riker (1987).

[4] Baron (2000) analyzed a screening model with transfers that shares many features of the Gilligan and Krehbiel (1987, 1989) and Krishna and Morgan (2001) models.

[5] In the equilibrium if the two agents propose different policies, the principal's beliefs jump discontinuously depending on the ordering and magnitudes of the two proposals. Battaglini (2002) developed a theory of robustness and demonstrated that this type of equilibrium is not robust to small probabilities that the informed agents are mistaken.

2 Environments

The results will be developed for relatively general environments and motivated by reference to the more specific models of Gilligan and Krehbiel (1989), Krishna and Morgan (2001), and Battaglini (2002). Consider settings with a principal ρ and a set $N = \{1, 2, \ldots, i, \ldots, n\}$ of agents. The state of the world is $\omega \in \Omega \subset \mathbb{R}^z$. We say $\omega \geq \omega'$ if each coordinate of ω is at least as large as that of ω'. The common prior belief over ω is represented by a distribution function $F(\omega)$ with support Ω.

The principal is to select a policy $p \in X \subset \mathbb{R}^k$. The principal and agents have von Neumann-Morgenstern utility functions over the state and policy, which we express by $u_j(\cdot, \cdot) : X \times \Omega \to \mathbb{R}^1$ for $j \in N \cup \{\rho\}$. Given the prior belief the expected utility of j from policy p is $\int u_j(p, \omega)\mathrm{d}F(\omega)$. We assume that for any $\omega \in \Omega$, the set $P_\rho(\omega)$ of full-information optimal policies for the principal is non-empty, where

$$P_\rho(\omega) := \arg\max_{p \in X} u_\rho(p, \omega).$$

To denote the states that lead to a particular optimal policy p, we write $P_\rho^{-1}(p) = \{\omega \in \Omega \mid p \in P_\rho(\omega)\}$, which is non-empty if $p \in \cup_{\omega \in \Omega} P_\rho(\omega)$.

In the information structure each agent $i \in N$ observes private information $\sigma_i = \sigma_i(\omega) \in \Sigma_i \subseteq \mathbb{R}^{z_i}, z_i \leq z$, which is a measurable and onto function $\sigma_i(\cdot) : \Omega \to \Sigma_i$ where $\Sigma_i = \{\sigma_i \mid \sigma_i = \sigma_i(\omega), \omega \in \Omega\}$. The principal has no private information. We use the notation $\Sigma := \times_{i \in N} \Sigma_i$, so $\sigma = \sigma(\omega) = (\sigma_1(\omega), \ldots, \sigma_n(\omega)) \in \Sigma$. The information structure is composed of the prior beliefs and the mappings $\sigma_i(\cdot), \forall i \in N$.

Each agent $i \in N$ sends a message $m_i = (m_{i1}, \ldots, m_{iz_i}) \in \Sigma_i$, and $m = (m_1, \ldots, m_n) \in \Sigma$ denotes a message profile. A message profile m that omits agent i's message is denoted m_{-i}. We use analogous notation σ_{-i} for signals. An environment is a tuple

$$\left\langle N, \rho, X, \Omega, F(\cdot), \left\{u_j(\cdot, \cdot)\right\}_{j \in N \cup \{\rho\}}, \left\{\Sigma_{j,} \sigma_j(\cdot)\right\}_{j \in N} \right\rangle.$$

The models of Gilligan and Krehbiel (1987, 1989) and Krishna and Morgan (2001) have $n = 2$, $\Omega = [0, 1]$, $F(\omega) = \omega$ on Ω, $X = \mathbb{R}$, $u_j\ (p, \omega) = -(p + \omega - y_j)^2$ for ideal points y_j, and private information $\sigma_i(\omega) = \omega$. In these models, each agent's private information provides perfect information about the state.

We consider two types of games. In a *signaling game*, each agent $i \in N$ observes $\sigma_i(\omega)$, and then the agents simultaneously send messages $m_i \in \Sigma_i$. The principal then selects a policy $p \in X$. A pure strategy perfect Bayesian equilibrium (PBE) to the signaling game is an n-tuple of message strategies

$$m_i(\cdot) \colon \Sigma_i \to \Sigma_i, \quad i = 1, \ldots, n,$$

a policy function

$$p(\cdot) \colon \Sigma \to X,$$

and a belief mapping

$$\mu(\cdot \mid \cdot) \colon \Sigma \to \triangle(\Omega),$$

where $\triangle(\Omega)$ denotes the space of probability measures on Ω for which (1) the message strategies are simultaneous best responses given the policy function [that is they form a Bayesian Nash equilibrium to the simultaneous move game of sending messages that influence policy through the policy function $p(\cdot)$], (2) the belief mapping satisfies Bayes' rule when it applies, and (3) the policy function is sequentially rational given the belief mapping (that is for every possible message profile m the policy function $p(m)$ selects a policy that optimizes the principal's expected utility $\int u_j(p, \omega) d\mu(\omega \mid m)$. A PBE is said to be *fully-revealing* if for almost every $\omega \in \Omega$, $\mu(\omega \mid m(\omega)) = 1$. In any fully-revealing PBE, sequential rationality requires that for almost every $\omega \in \Omega$,

$$p(m_1(\sigma_1(\omega)), \ldots, m_n(\sigma_n(\omega))) \in P_\rho(\omega).$$

In a *screening game* each agent $i \in N$ observes $\sigma_i(\omega)$, and then the principal announces and commits to a *mechanism*

$$g(\cdot) \colon \Sigma \to X.$$

Following this announcement each agent simultaneously sends a message $m_i \in \Sigma_i$, and the policy is given by $g(\cdot)$.[6] In the screening game a revelation principle holds so we consider only direct mechanisms. A mechanism is *truthful* if for every $i \in N$ $m_i(\sigma_i) = \sigma_i$ is a simultaneous best response given the mechanism (i.e., truthful strategies form a Bayesian Nash equilibria given the mechanism). A truthful mechanism is *full-information optimal* (for the principal) if for a.e. $\omega \in \Omega$, $g(m_1(\sigma_1(\omega)), \ldots, m_n(\sigma_n(\omega))) \in P_\rho(\omega)$. In the remainder of the analysis we consider only full-information optimal truthful mechanisms, which we refer to as full-information optimal mechanisms.

Since the interest here is in equilibria that fully reveal the state to the principal, the information structure must be such that ω can be inferred from the private information σ. To ensure this, the function $\sigma(\cdot)$ is assumed to be one-to-one. The models considered by Gilligan and Krehbiel (1987, 1989), Krishna and Morgan (2001), and Battaglini (2002) satisfy this condition, since all agents observe the true state.

Invertibility of $\sigma(\cdot)$ is the minimal property of the private information that allows full-revelation of the state. In particular, if the agents are truthful, the principal can invert the message profile to infer ω. This condition does not require that the individual $\sigma_i(\cdot)$ functions reveal the state. An information structure that allows full revelation but does not have one-to-one individual signal functions is: $\Omega = \mathbb{R}^2$ with generic element $\omega = (\omega^1, \omega^2)$, $n = 2$, $\sigma_1(\omega) = \omega^1$, $\sigma_2(\omega) = \omega^2$.

We term a message profile *m supportable* if there is an $\omega' \in \Omega$ for which

$$(m_1, \ldots, m_n) = \left(\sigma_1(\omega'), \ldots, \sigma_n(\omega')\right).$$

We use the term *unsupportable* if m is not supportable.

To relate the equilibria of the signaling and corresponding mechanism design games, we consider environments in which the principal's policy choice following an unsupportable message profile can be justified by some beliefs. As shown in the next section, justifiable beliefs are necessary for the equivalence between

[6] We consider only mechanisms without transfers to keep the comparison with signaling games clear.

fully-revealing signaling equilibria and full-information optimal mechanisms in the corresponding screening game.

Definition 1 *A mechanism* $g(\cdot)$ *is said to be belief-justifiable if following every* m *that is unsupportable, there is some conditional distribution* $F(\cdot; m)$ *on* Ω *for which* $g(m) \in \arg\max_{x \in X} \int_{\Omega} u_{\rho}(x, \omega) \mathrm{d}F(\omega; m)$.

Following an unsupportable message profile this condition requires that the mechanism selects a policy that is the best response for the principal given some belief about the state. The strategy of this paper is to use the property of belief-justifiability to identify full-information optimal mechanisms $g(\cdot)$ that are sequentially rational both on and off the equilibrium path and hence are best-responses in a PBE of the corresponding signaling game. In the mechanism design game a full-information optimal policy is related to sequential rationality on the equilibrium path, and belief-justifiability is related to sequential rationality off the path. When a mechanism is full-information optimal following every supportable m, the mechanism chooses a policy that is sequentially rational for some consistent belief. When a full-information optimal mechanism is belief-justifiable, following every unsupportable m, there is some belief for which the policy choice is sequentially rational. If a direct mechanism fails to satisfy either of the conditions (full-information optimal or belief justifiable), there will be no PBE in the corresponding signaling game that yields the same policy function. This may also be interpreted as a two-stage mechanism in which in the first stage the principal commits to beliefs and in the second stage chooses a policy. In the Gilligan and Krehbiel (1987, 1989), Krishna and Morgan (2001), and Battaglini (2002) models, it is sufficient to focus on off-the-equilibrium-path beliefs in the PBE that are concentrated on a single state. As shown in the penultimate section the beliefs that support the full-information optimal mechanism can in some environments also be specified in a simple manner.

3 Equivalence between full-information optimal mechanisms and fully-revealing PBE

The following straightforward propositions identify an equivalence between the existence of truthful PBE of the signaling game and the existence of full-information optimal mechanisms in the screening game.

Proposition 1 *If there exists a fully-revealing PBE of the signaling game, there exists a belief-justifiable, full-information optimal mechanism* $g(\cdot)$ *in the screening game.*

Proof Assume there exists a fully-revealing PBE

$$\langle \{m_i(\cdot)\}_{i \in N}, p(\cdot), \mu(\cdot \mid \cdot) \rangle$$

to the signaling game. Since the $m_i(\cdot)$ are simultaneous best responses, the mechanism $g(\cdot) = p \circ m$ is a truthful direct mechanism. Sequential rationality implies that for every ω, $p(m(\sigma(\omega))) \in P_{\rho}(\omega)$, so the mechanism $g(\cdot)$ satisfies the optimality condition in the definition of full-information optimal. To verify that this mechanism is belief-justifiable, note that since $p(\cdot)$ is sequentially rational for the principal it must be the case that for every m (and therefore every unsupportable m) there is a belief $F(\cdot; m)$ with support Ω for which $p(m) \in \arg\max_{x \in X} \int_{\Omega} u_{\rho}(x, \omega) \mathrm{d}F(\omega; m)$. Thus, a full-information optimal mechanism exists that is belief-justifiable. □

Although this result is straight-forward and nearly obvious, its significance for this paper is found in its contra-positive.

Corollary 1 *If in a screening model, there is no full-information optimal mechanism that is belief-justifiable, there is no fully-revealing equilibrium in the corresponding signaling model.*

Corollary 1 is useful because it shows that while belief justifiability might be a strong requirement, it is *necessary* if we are to construct PBE in a signaling game from a full-information optimal mechanism in a screening game. The corollary allows us to show that certain environments do not possess fully-revealing equilibria in the signaling game by analyzing only the screening game. The following result establishes the converse of Proposition 1 and motivates the presentation of sufficient conditions for belief-justifiable mechanisms in the next section.

Proposition 2 *If there exists a belief-justifiable, full-information optimal mechanism $g(\cdot)$ in the screening game, there exists a fully-revealing PBE of the signaling game.*

Proof Assume that $g(\cdot)$ is a belief-justifiable, full-information optimal mechanism. Since $g(\cdot)$ is belief-justifiable, if m is unsupportable there are beliefs $F(\cdot; m)$ such that

$$g(m) \in \arg\max_{x \in X} \int_{\Omega} u_{\rho}(x, \omega) \mathrm{d}F(\omega; m).$$

On the equilibrium path (following a supportable m), the beliefs are determined by assigning probability 1 to the set $\{\omega : \sigma(\omega) = m\}$. Since $g(\cdot)$ is a truthful direct mechanism and the mapping $\sigma(\cdot)$ is one-to-one, this set is a singleton. Since $g(\cdot)$ is truthful, these beliefs satisfy Bayes' rule following a supportable m (which is exactly the set of message profiles for which Bayes' rule applies). Accordingly, the beliefs described are well-defined. A fully-revealing PBE is then

$$\mu(\omega \mid m) = \begin{cases} 1 \quad \text{if } m \text{ is supportable for some } \omega' \in \Omega \text{ and } \omega \geq \omega' \\ F(\omega; m) \text{ if } m \text{ is unsupportable} \\ 0 \text{ otherwise} \end{cases}$$

and

$$m_i(\sigma_i) = \sigma_i.$$

The strategy $m_i(\sigma_i) = \sigma_i$ is a best response to $m_j(\cdot)$, $j \in N \backslash i$ given $p(\cdot)$ because the mechanism $g(\cdot)$ is truthful and $p(\cdot) = g(\cdot)$. The strategy $p(\cdot)$ is sequentially rational given $\mu(\cdot \mid \cdot)$ because $\mu(\omega \mid m)$ is concentrated at an element of $P_{\rho}^{-1}(g(m))$ for every supportable m and $F(\cdot; m)$ is a belief that makes $g(m)$ the best response for the principal following an unsupportable m. □

4 Existence of full-information optimal mechanisms

Given the equivalence results of the preceding section, this section presents two sufficient conditions for the existence of a belief-justifiable full-information optimal mechanism. These two conditions are not tight and suggest that weaker conditions might be identified. The first is a strong condition on the relevance of individual information. If when any single agent deviates from truthful messages the principal still can infer the true state, full-information optimal policy functions can be constructed that are unaffected by any single deviation. Accordingly, such mechanisms are truthful. The second condition pertains to settings in which the principal can determine whether a single agent lied. The condition involves a policy, on the one hand, optimal for the principal under some off-the-equilibrium-path beliefs, and on the other hand, sufficiently unattractive to the senders in all states. In cases in which no such punishment exists or the receiver cannot determine the state following a profile of messages that is not truthful, weaker conditions may exist but the analysis must involve policy functions that base the punishment on the specific profile of messages observed. In both cases, the results require only the existence of a truthful equilibrium given the mechanism. The mechanisms may also admit equilibria in which agents are not truthful.

4.1 Sufficient information structures

One sufficient condition is an information structure that allows the principal to infer the true state when only one agent misrepresents her signal. This condition is satisfied, e.g., whenever three or more agents receive the same information. Recall that $\sigma_{-i}(\cdot)$ is a mapping from Ω onto $\times_{j\in N\backslash i}\Sigma_j$. Similarly, for any distinct $i, j \in N$ let $\sigma_{-ij}(\cdot)$ denote the mapping from Ω onto $\times_{k\in N\backslash\{i,j\}}\Sigma_k$.

Definition 2 *We say that the information structure is strongly non-exclusive if for any distinct $i, j \in N$ the mapping $\sigma_{-ij}(\cdot)$ is one-to-one.*

The primary consequence of this condition is that when $m = (\sigma_{-j}(\omega'), \alpha)$ for some $j \in N, \alpha \in \Sigma_j$ and $\omega' \in \Omega$, the set $\{\omega \in \Omega : \exists i \in N \text{ s.t. } m_{-i} = \sigma_{-i}(\omega)\}$ is equivalent to $\{\omega'\}$. A simple heuristic demonstrates this point. Following truthful messages by all agents except k, if the principal considers all the $n-2$ message profiles m_{-ij}, she will find that all unsupportable profiles include a message from k and that profiles that exclude k are supportable. This information is enough to convince the principal that if at least $n-1$ agents are truthful, agent k is not being truthful. Moreover, by inverting any of the profiles that exclude k the principal can identify ω.

This condition is vacuous unless there are at least three agents ($n \geq 3$). In the settings considered by Gilligan and Krehbiel (1987, 1989), Krishna and Morgan (2001), and Battaglini (2002), the information structure is strongly non-exclusive if there are three or more agents, each observing ω. Strong non-exclusivity applies to more general information structures. As an example of a strong non-exclusive information structure in which no two agents have perfectly correlated signals, consider $\Omega = [0, 1]^4$, six agents, and signals $\sigma_1(\omega) = (\omega^1, \omega^2)$; $\sigma_2(\omega) = (\omega^1, \omega^3)$; $\sigma_3(\omega) = (\omega^1, \omega^4)$; $\sigma_4(\omega) = (\omega^2, \omega^3)$; $\sigma_5(\omega) = (\omega^2, \omega^4)$; $\sigma_6(\omega) = (\omega^3, \omega^4)$. In this example, each coordinate of ω is observed by three agents. To see how this structure allows the principal to infer the signal $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ given a unilateral deviation from a truthful message, suppose the principal anticipates that the agents are sending truthful messages, but the second coordinate of m_2, the second coordinate of m_4 and the first coordinate of m_6 do not coincide. The principal then concludes that at least one of these agents is not being truthful. Since she has three statements about ω^3, if two of the statements coincide, then she can assume that the other message is not truthful.

Strong non-exclusivity is stronger than the condition of Non-exclusivity in information introduced by Postlewaite and Schmeidler (1986) and restated in our notation below in definition 4.[7] The relation between the conditions is straightforward. An information structure with n agents satisfies strong non-exclusivity if and only if for any subset of $n-1$ agents, the information structure satisfies Postlewaite and Schmeidler's non-exclusivity condition. With non-exclusivity, following truthful messages by $n-1$ agents, the principal can detect that one agent was not truthful, but she may not be able to identify ω. With strong non-exclusivity, the principal can also identify ω and which agent was not truthful.

The next result establishes the existence of full-information optimal mechanisms and fully-revealing PBE for strongly non-exclusive information structures regardless of agent and principal preferences. Strong non-exclusivity allows the principal to specify a full-information optimal mechanism without the use of punishments, which may be limited when utility is not transferable.

Proposition 3 *For a strongly non-exclusive information structure with $n > 2$,* (i) *a belief-justifiable full-information optimal mechanism $g(\cdot)$ exists and hence* (ii) *a fully-revealing PBE of the signaling game exists.*

Proof (i) Since $\sigma(\cdot)$ is one-to-one, following a supportable m a unique ω for which $m = \sigma(\omega)$ exists. Additionally, in a strongly non-exclusive information structure, if m is unsupportable and only one agent is not truthful, the set $\{\omega \in \Omega : m_j = \sigma_j(\omega) \text{ for all } j \in N\backslash i \text{ for some } i \in N\}$ contains only the true ω. Let $p_\rho(\cdot)$ be a single-valued selection from the correspondence $P_\rho(\omega)$. We now construct the mechanism:[8]

$$g(m) = \begin{cases} p_\rho(\omega) & \text{if } m \text{ is supportable and } \omega = \sigma^{-1}(m) \\ p_\rho(\omega') & \text{if } m \text{ is unsupportable and } \exists i \in N \text{ s.t. } \omega' = \sigma_{-i}^{-1}(m_{-i}) \\ p_\rho(\omega'') & \text{otherwise, for an arbitrary } \omega'' \in \Omega. \end{cases}$$

Since $g(\cdot)$ is unaffected by a single agent's deviation from a truthful profile, if every agent except i is truthful, $m_i(\sigma_i) = \sigma_i$ is a best response for i. Thus, the mechanism is truthful. Since $g(m) \in P_\rho(\omega)$ following a supportable $m = \sigma(\omega)$, the mechanism is full-information optimal. Since $g(m) \in P_\rho(\omega'')$ for some $\omega'' \in \Omega$ following an unsupportable m, the mechanism is belief-justifiable.

(ii) This follows from Proposition 2. □

[7] See also Palfrey and Srivastava (1987) for the use of this condition in implementation.

[8] This mechanism admits other equilibria that are not truthful. For instance, if all agents use a constant message function (which is supportable), then no agent will have a unilateral incentive to deviate.

With three or more agents who all observe the true state, a full-information optimal mechanism trivially exists, since $g(\cdot)$ is unaffected by a deviation by a single agent. Proposition 3 extends this to a class of information structures in which agents observe different signals and only $n-2$ signals are needed to identify the state. To exhibit such a mechanism, consider the example following the definition of strong non-exclusivity. Letting $X = \mathbb{R}^4$ and $u_\rho(p) = -\sum_{i=1}^{4}(\omega^i + p^i)^2$, the mechanism is

$$g(m) = \begin{cases} -m \text{ if } m = \sigma(\omega) \text{ or } m_{-i} = \sigma_{-i}(\omega) & \text{for some } i \in N \text{ and } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

No agent has an incentive to deviate from a truthful message provided the other agents are believed to be sending truthful messages.

Combining Proposition 3 with Battaglini's (2002) results for his multidimensional extension of the Gilligan and Krehbiel (1987, 1989) models yields the following characterizations. If there is only one agent, no fully-revealing PBE or full-information optimal mechanism (without transfers) exists. If there are two senders and $X, \Omega \subset \mathbb{R}^1$, there are fully-revealing PBE and full-information optimal mechanisms if the conflict of interest between the agents is not too large relative to the uncertainty.[9] If the dimensionality of X and Ω is two or more, generically there are fully-revealing PBE and full-information optimal mechanisms with two or more agents. If there are three or more agents, then for any dimensionality, Proposition 3 implies there are always fully-revealing PBE and full-information optimal mechanisms.[10]

4.2 Punishments

The condition of strong non-exclusivity is sufficient to allow the principal to infer the true state when any $n-1$ agents are truthful, so a belief-justifiable full-information optimal mechanism exists. Another condition sufficient for the existence of such a mechanism relies on punishments for any deviation from truthful messages. A punishment must be disadvantageous to any deviator, but without transfers it may be impossible to punish only the deviator. First, if strong non-exclusivity is not satisfied, it may be impossible to determine which agent deviated. Second, it may be impossible to choose a policy that punishes the deviator

[9] See Battaglini's (2002) Proposition 1 where the condition $|x_1| + |x_2| > W$ is shown to be necessary and sufficient for the non-existence of fully-revealing PBE, where x_i is the ideal point of agent i and $\Omega = [-W, W]$.

[10] The (belief-justifiable) full-information optimal mechanisms and fully-revealing PBE that attain in strongly non-exclusive information environments need not be coalition-proof. For example, consider three agents observing ω. If the agents and principal have strictly concave preferences over the outcome $p + \omega$ and if the principal does not share the ideal point of any agent then at least two agents prefer to bias policy in one direction. While Battaglini's equilibria of the two-agent game are coalition-proof, his notion of coalition-proofness does not allow the agents to reach a binding agreement on the messages they send to the principal. Since the agents' preferences differ from each other on both dimensions, there is no self-enforcing agreement that benefits the agents. Battaglini's equilibria are not collusion-proof in the sense used by Laffont and Martimort (1997), who assume that the agents can make an enforceable collusive agreement. Such an agreement forces the principal to make a costly adjustment to the mechanism to negate the incentives to collude.

without attracting a deviation by another agent. Instead of attempting to identify punishments specific to each particular unsupportable profile of messages, conditions are presented such that all agents are punished when any agent deviates from a truthful message profile. This requires that (a) the information structure allows a deviation to be detected and (b) there is a policy that sufficiently punishes all agents in every state so that the threat of this punishment discourages a deviation. The following two definitions correspond to (a) and (b), respectively.

Definition 3 *We say the information structure satisfies non-exclusivity if for every* $i \in N$ *the mapping* $\sigma_{-i}(\cdot)$ *is one-to-one.*

As noted, this condition corresponds to Postlewaite and Schmeidler's Non-exclusivity of information. An environment satisfies non-exclusivity if when the principal believes the agents are sending truthful messages and exactly one agent does not send a truthful message, the principal can infer that at least one agent was not truthful. That is, given the belief that agents are sending truthful messages, the principal knows that there is no state consistent with the message profile and hence knows they are to be punished. This does not require that the principal can infer ω or the identity of the agent that lied following a profile with exactly one non-truthful message. Note that strong non-exclusivity implies non-exclusivity.

To illustrate the concept of non-exclusivity, consider an example with three agents and $\omega \in \Omega = [0, 1]^3$. If $\sigma_1(\omega) = (\omega^1, \omega^2)$, $\sigma_2(\omega) = (\omega^2, \omega^3)$ and $\sigma_3(\omega) = (\omega^1, \omega^3)$, then non-exclusivity is satisfied. In this case, however, following a non-truthful message about ω^1 by player 1, the principal cannot determine whether player 1 or player 3 is lying. Moreover, she cannot determine the true value of ω^1; so the information structure is not strongly non-exclusive. If $\sigma_1(\omega) = (\omega^1)$, $\sigma_2(\omega) = (\omega^2)$ and $\sigma_3(\omega) = (\omega^3)$, then non-exclusivity is not satisfied, since a deviation by any player cannot be detected. In a two-agent model, non-exclusivity requires that both agents' signal mappings $\sigma_i(\cdot)$, $i = 1, 2$, are one-to-one. Non-exclusivity is related to the statistical properties of the private information mappings $\sigma_i(\cdot)$. For example, non-exclusivity cannot be satisfied if all the individual mappings are independent, but it does not require perfect correlation among individual signals as is assumed in the two-sender models of Gilligan and Krehbiel (1987, 1989), Krishna and Morgan (2001), and Battaglini (2002).

The following condition on punishment is used.

Definition 4 *We say an environment satisfies* **punishability** *if there exists an* $r^* \in X$ *and a function* $p^*(\cdot) : \Omega \to X$ *such that* (1) *for some belief* $\lambda(\cdot)$ *on* Ω, $r^* \in \arg\max_{x \in X} \int_\Omega u_\rho(x, \omega) d\lambda(\omega)$, (2) $p^*(\omega) \in P_\rho(\omega)$ *for a.e.* $\omega \in \Omega$, *and* (3) *for a.e.* $\omega \in \Omega$ *the following is true for all* $i \in N$

$$\int_\Omega u_i(p^*(\omega'), \omega') \mathrm{d}\eta_i(\omega' \mid \sigma_i(\omega)) \geq \int_\Omega u_i(r^*, \omega') \mathrm{d}\eta_i(\omega' \mid \sigma_i(\omega)),$$

where the probability measure $\eta_i(\cdot \mid \cdot)$ *satisfies Bayes' rule conditional on the signal* $\sigma_i = \sigma_i(\omega)$.[11]

[11] The preferences in the inequality are *ex interim*, and the measure $\eta_i(\cdot \mid \sigma_i(\omega))$ is the *ex interim* belief. Note that the principal's beliefs $\lambda(\cdot)$ are not required to satisfy Bayes' rule.

This is never satisfied in the environment studied by Battaglini (2002), but it may be satisfied in other environments. Environments satisfy punishability if for some selection of policies $p^*(\omega)$ that are optimal given the true state there is a punishment policy r^* that the principal would select given some beliefs and is sufficiently undesirable that each agent prefers being truthful and attaining $p^*(\omega)$ to attaining r^*.

A simple example of an environment that satisfies punishability involves a principal choosing a policy based on information provided by two interest groups with opposing preferences. Regardless of the state each interest group prefers that the principal choose a policy different from the status quo, but they differ on which policy to choose. The principal prefers a policy that is appropriate for the state, and in one state she is indifferent between the status quo and a compromise policy. This simple example is modelled as follows: $X = \{x_1, x_2, x_3, 0\}$, $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $N = \{1, 2\}$, $\sigma_i(\omega) = \omega \in \Omega$. The principal prefers x_1 when the state is ω_1, x_2 when the state is ω_2, and between x_3 and 0 when the state is ω_3 is indifferent but prefers either of these policies to the other alternatives. So, 0 could be the status quo and x_3 a compromise policy. Agent 1 prefers x_1 to x_3 to x_2 to 0 in any state. Agent 2 prefers x_2 to x_3 to x_1 to 0 in any state. The principal can use the following mechanism

$$g(m) = \begin{cases} x_1 & \text{if } m_1 = m_2 = \omega_1 \\ x_2 & \text{if } m_1 = m_2 = \omega_2 \\ x_3 & \text{if } m_1 = m_2 = \omega_3 \\ 0 & \text{otherwise} \end{cases}$$

This mechanism can be supported as a sequentially-rational policy function in a fully-revealing PBE of the signaling game by allowing the principal to form the belief that $\omega = \omega_3$ with probability 1 following any unsupportable m, in which case the principal is willing to choose 0. Given that agent 1 is truthful, any non-truthful message by agent 2 will result in the policy 0 which is least desirable for that agent.

Given an environment satisfying non-exclusivity and punishability, the construction of a full-information optimal mechanism is straightforward.

Proposition 4 *If an environment satisfies non-exclusivity and punishability,* (i) *there is a belief-justifiable full-information optimal mechanism, and hence* (ii) *a fully-revealing PBE of the signaling game exists.*

Proof (i) If punishability is satisfied, a function $p^*(\cdot)$ and policy r^* satisfying Definition 4 exist. A full-information optimal mechanism is:[12]

$$g(m) = \begin{cases} p^*(\omega') & \text{if } m \text{ is supportable and } m = \sigma(\omega') \\ r^* & \text{if} m \text{ is unsupportable} \end{cases}$$

Since the mapping $\sigma(\cdot)$ is one-to-one, the ω' in the definition of $g(\cdot)$ is unique. If every agent other than i is truthful and agent i is truthful, his expected utility is given by the left-side of the inequality in Definition 4. If only agent i is not truthful, by non-exclusivity, the principal knows that the message profile is

[12] Note that if the condition that r^* is optimal for some belief were dropped from the definition of punishability the mechanism $g(\cdot)$ would still be full-information optimal. This condition is used below to establish belief-justifiability of the mechanism.

unsupportable. The mechanism then selects r^*, giving i expected utility equal to the right-side of the inequality. Thus, a truthful message is a best response for i when every other agent is truthful, establishing that the mechanism is truthful. Since $p^*(\omega) \in P_\rho(\omega)$, the mechanism is full-information optimal. If an agent deviates and the message profile is unsupportable, punishability implies that there is a belief $\lambda(\cdot)$ such that r^* is optimal for the principal, establishing that the mechanism is belief-justifiable.

(ii) The result follows from Proposition 2. □

Informally, Proposition 4 states that if the principal can detect when exactly one agent lies, the availability of a policy that is belief-justifiable and undesirable to the agents is sufficient for the existence of a full-information optimal mechanism. In particular, environments more specific or focused types of punishment may be possible. For example, the equilibrium Krishna and Morgan (2001) characterize in their Proposition 1 and the class of equilibria Battaglini (2002) uses to prove his Proposition 1 are not similar to the equilibrium identified in Proposition 4 here. In their equilibria the policy function uses a punishment that is responsive to the messages and only a particular agent is punished, whereas the mechanism used in the construction in Proposition 4 calls for the same punishment following any unsupportable m. Punishability is not satisfied in the Krishna and Morgan (2001) and Battaglini (2002) models. To see this, consider (in slightly different notation than Battaglini's) the state space $\Omega = [-W, W]$ with the principal's utility function $u_\rho(p,\omega) = -(p+\omega)^2$ and agent j's utility functions $u_j(p,\omega) = -(y_j - p - \omega)^2$ with $-W < y_1 < 0 < y_2 < W$. Any belief-justifiable policy r must be contained in the set $[-W, W]$. But for each agent any policy in this set is optimal for some $\omega \in [-W, W]$. This means that no belief-justifiable punishment r is sufficiently undesirable to all the agents to make them always prefer the principal's ideal outcome 0 to the outcome $r + \omega$ resulting from the policy r.

It should be noted that the fully-revealing PBE that attain when punishability holds need not be robust in the sense that Battaglini (2002) discusses.[13] The equilibria may involve beliefs that are highly discontinuous in the messages. In contrast, in a non-exclusive information environment, there will always be robust fully-revealing PBE. These equilibria involve beliefs that are unaffected by a single agent's deviation and are, therefore, continuous in m near the equilibrium path.

As a final example, consider two interest groups, $N = \{1, 2\}$ and a policy space $X = \{x_1, x_2, x_3\}$ with $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\sigma_i(\omega) = \omega$ for $i \in N$. Suppose the state and policy contingent utilities of the principal and agents 1 and 2 are given by the respective entries in the matrix

$x \backslash \omega$	ω_1	ω_2	ω_3
x_1	2, 1, 2	1, 1, 2	1, 2, 1
x_2	1, 2, 1	2, 2, 1	1, 1, 2
x_3	a, 0, 0	b, 0, 0	c, 0, 0

If a, b, c are all less than 1, no belief-justifiable mechanism can ever select x_3. This means that in the search for a belief-justifiable full-information optimal

[13] His notion of robustness involves considering a perturbed game in which each agent receives incorrect information (with support Σ_i) with probability ε_i. A fully-revealing PBE is robust if the principal's beliefs are the limit of beliefs in some sequence of perturbed games where $\max_{i \in N} \varepsilon_i$ tends to 0.

mechanism, we can focus on mappings $g(\cdot, \cdot)$ with image $\{x_1, x_2\}$. Now, given the principal's preferences any truthful full-information optimal mechanism must satisfy $g(\omega_1, \omega_1) = x_1$, $g(\omega_2, \omega_2) = x_2$ and $g(\omega_3, \omega_3) \in \{x_1, x_2\}$. For a truthful message to be the best response for agent 1 (2) when agent 2 (1) is truthful, it must be the case that $g(\omega_2, \omega_1) \neq x_2$ ($g(\omega_2, \omega_1) \neq x_1$). But since belief-justifiability requires that only x_1 or x_2 is chosen, no belief-justifiable full-information optimal mechanisms exist in this environment. Applying Corollary 1 we see that there is also no fully-revealing PBE of the corresponding signaling game. This conclusion follows without having to consider beliefs. Instead, the argument focuses exclusively on interpreting the requirements imposed by belief-justifiability and the optimality conditions on the policy function. In contrast, there are full-information optimal (but not belief-justifiable) mechanisms. One example is

$$g(m_1, m_2) = \begin{cases} x_1 & \text{if } m_1 = m_2 = \omega_1 \\ x_2 & \text{if } m_1 = m_2 = \omega_2 \\ x_3 & \text{otherwise.} \end{cases}$$

However, if $c \geq 1$, this mechanism would also be belief-justifiable and by proposition 2 there would be a fully-revealing PBE in the signaling game.

5 Discussion

When modelling collective choice with asymmetric information, the modeler can choose among alternative models or game forms. In many settings, empirical evidence informs this choice. In other settings it is necessary to consider the universe of possible institutions. The politics literature has relied on signaling models involving equilibrium requirements (sequential rationality) on the principal's actions. In contrast, in screening models the principal is able to commit to a policy function that need not be sequentially rational. For a class of environments that include those of Gilligan and Krehbiel (1987, 1989), Krishna and Morgan (2001), and Battaglini (2002) and yield fully-revealing equilibria, these two approaches yield the same equilibrium outcomes. This result can be interpreted in two ways. If one is content to seek PBE in signaling games, a useful intermediate step can be to solve the sometimes simpler screening problem. Alternatively, if one is bothered by the commitment assumption of the screening game, simply focusing on the set of PBE of the signaling game should also be bothersome. That is, without appeal to refinements of off the equilibrium path beliefs, equilibria to signaling games may in effect involve commitment masked behind highly specific off-the-equilibrium-path beliefs. The PBE concept allows the principal to commit to these off-the-equilibrium-path beliefs, and given sequential rationality these beliefs in effect commit the principal to a policy.

The sufficient conditions in Sect. 4, while far from necessary for particular environments, are sometimes reasonable. In the simplest case in which agents all receive the same signal, it is only with one or two agents that it may be difficult for the principal to learn the true state. In the settings in which the signaling models have been tested (Krehbiel, 1991), the assumption of only one or two agents may be unrealistic. With more agents, the prediction of fully-revealing equilibria regardless of the dimensionality of the choice set or the information set and the degree of preference divergence, may be reasonable provided that the agents have access to the same information. That assumption, however, may not be satisfied in other strategic settings.

The assumption of punishability, while incompatible with the basic structure of the signaling models cited above, may also be satisfied in applied settings. Returning to the penultimate example of a principal that has to make a policy choice and requests information from two interest groups, a feasible punishment can exist. If the two interest groups provide incompatible information, the principal can simply choose as a punishment the status quo which for the principal is as good as the compromise alternative. This punishment is sequentially rational in a PBE given beliefs that no alternative to the status quo is desirable when the agents send incompatible messages.

This paper has explored the implications of the considerable flexibility in belief formation afforded by the equilibrium concepts used in signaling models with a continuum of states. Moreover, it demonstrates that in some environments, the assumption of commitment in screening models is no more problematic than the use of standard equilibrium concepts in signaling games that allow considerable flexibility in the specification of off-the-equilibrium-path beliefs. The task of determining which model and equilibrium concept to use in applications where the equilibria are not equivalent remains for the researcher.

Acknowledgements This research has been supported by NSF Grant No. SES-0111729. We thank David Ahn and Thomas Palfrey for their comments.

References

Austen-Smith D, Riker W (1987) Asymmetric information and the coherence of legislation. Ame Polit Sci Rev, 81(3):897–918

Baron DP (2000) Legislative organization with informational committees. Am J Polit Sci, 44:506–522

Battaglini M (2002) Multiple referrals and multidimensional cheap talk. Econometrica, 70:1379–1401

Crawford VP, Sobel J (1982) Strategic information transmission. Econometrica, 50(6):1431–1451

Cremer J, McLean RP (1985) Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. Econometrica, 53(2):345-361

Cremer J, McLean RP (1988) Full extraction of the surplus in Bayesian dominant strategy auctions. Econometrica, 56(6):1247–1257

Durrett R (1996) Probability Theory and Examples. 2nd edn. Duxbury Press, Belmont

Gilligan TW, Krehbiel K (1987) Collective decision-making and standing committees: an informational rationale for restrictive amendment procedures. J Law, Econ Organ, 3(2):287–335

Gilligan TW, Krehbiel K (1989) Asymmetric information and legislative rules with a heterogenous committee. Am J Polit Sci 33(2):459–490

Krehbiel K (1991) Information and legislative organization. University of Michigan Press, Ann Arbor

Krehbiel K (2001) Plausibility of signals by a heterogenous committee. Ame Polit Sci Rev 95(2):453-457

Krishna V, Morgan J (2001) Asymmetric information and legislative rules: some amendments. Am Polit Sci Rev 95(2):435-452

Laffont J-J, Martimort D (1997) Collusion under asymmetric information. Econometrica 65:875–911

McAfee RP, Reny PJ (1992) Correlated information and mechanism design. Econometrica 60(2):395–421

Melumad ND, Shibano T (1991) Communication in settings with no transfers. RAND J Econ 22(2):173–197

Palfrey TR, Srivastava S (1987) On bayesian implementable allocations. Rev Econ Stud 54:193–208

Postlewaite A, Schmeidler D (1986) Implementation in differential information economies. J Econ Theory 39:14–33