IMPLEMENTATION VIA INFORMATION DESIGN IN BINARY-ACTION SUPERMODULAR GAMES

STEPHEN MORRIS, DAISUKE OYAMA, AND SATORU TAKAHASHI

ABSTRACT. What outcomes can be implemented by the choice of an information structure in binary-action supermodular games? An outcome is partially implementable if it satisfies obedience (Bergemann and Morris (2016)). We characterize when an outcome is smallest equilibrium implementable (induced by the smallest equilibrium) and fully implementable (induced by all equilibria). Smallest equilibrium implementation requires a stronger sequential obedience condition: there is a stochastic ordering of players under which players are prepared to switch to the high action even if they think only those before them will switch. Full implementation requires sequential obedience in both directions. Our characterization of smallest equilibrium implementation can be used to solve the information design problem with adversarial equilibrium selection.

(S. Morris) Department of Economics, Massachusetts Institute of Technology
(D. Oyama) Faculty of Economics, University of Tokyo
(S. Takahashi) Department of Economics, National University of Singapore
E-mail addresses: semorris@mit.edu, oyama@e.u-tokyo.ac.jp, ecsst@nus.edu.sg.
Date: November 12, 2020.

Keywords. Information design, supermodular game, smallest equilibrium implementation, sequential obedience, potential game.

We are grateful for the comments from Roberto Corrao and Junjie Zhou as well as those from seminar participants at the “Current Topics in Microeconomics Theory” conference at the City University of Hong Kong, the Theory and Finance Workshop on Coordination and Information Design at the People Bank of China School of Finance at Tsinghua University, the 2019 Stony Brook International Conference in Game Theory, the Northwestern Workshop on Computer Science and Economics, the 12th Econometric Society World Congress, and the One World Mathematical Game Theory Seminar, and at Waseda, Keio, Tsukuba, Yokohama National, Kobe, Osaka, Singapore Management, Harvard/MIT, Caltech, Kansai, Seoul National, Hitotsubashi, Cornell, and Columbia Universities. Stephen Morris gratefully acknowledges financial support from NSF Grant SES-2001208. Daisuke Oyama gratefully acknowledges financial support from JSPS KAKENHI Grants 18KK0359 and 19K01556.
1. Introduction

Consider an information designer who can choose the information structure for players in a game but cannot control what actions the players choose. The designer is interested in the induced joint distribution over actions and states, which we call an outcome. What outcomes can be implemented by information design?

This problem has been studied in recent years under the classical partial implementation assumption that the designer can also pick the equilibrium played. It is without loss of generality to restrict attention to direct mechanisms, where players are simply given an action recommendation by the information designer. An outcome can be partially implemented if and only if it satisfies an obedience constraint, i.e., the requirement that players have an incentive to follow the designer’s recommendation. This is equivalent to saying that the outcome is an incomplete information version of correlated equilibrium; Bergemann and Morris (2016) call the relevant version a Bayes correlated equilibrium.¹

We study how the answer to this question changes if we are interested in two more demanding notions of implementation: smallest equilibrium implementation and full implementation. We address these questions in the context of binary-action supermodular (BAS) games, where a smallest equilibrium will always exist. Smallest equilibrium implementation requires that the outcome be induced in the smallest equilibrium under the chosen information structure. Full implementation requires that the outcome be induced in all equilibria under the chosen information structure.

Our main result is a characterization of smallest equilibrium implementability. The characterization is closely analogous to the obedience characterization of partial implementation. In particular, it corresponds to a finite system of linear inequalities. The more demanding criterion of smallest equilibrium implementation gives rise to a more demanding sequential obedience constraint. Sequential obedience requires that it be possible for the information designer to choose (perhaps randomly conditioning on the state) an ordering of players in which players are advised to play the high action in such a

¹Bergemann and Morris (2019) provide an overview of a now large literature building on this observation. What we are calling the information design problem is a many player generalization of the Bayesian persuasion problem described by Kamenica and Gentzkow (2011) (see Kamenica (2019) for a survey of this literature). Bergemann and Morris (2013, 2016) characterized the implementable outcomes in the many player case and noted information design applications; Taneva (2019) suggested the terminology “information design”; this corresponds to the strand of mechanism design that Myerson (1991) labelled “communication in games”, with the twist that the designer is able to deliver information to the players without having to elicit it.
way that they are strictly willing to follow the recommendation even if they only expect players who received the recommendation before them to choose the high action.

To see why sequential obedience is necessary, suppose that an outcome is smallest equilibrium implementable. Then there exists an information structure where this outcome is played in the smallest equilibrium. Recall (from, e.g., Milgrom and Roberts (1990)) that it is a standard property of supermodular games that the smallest equilibrium can be reached by a myopic best response sequence, where we start with all types of all players choosing the low action and sequentially switch players’ types to the high action when it is a strict best response to do so (where the order of switches does not matter). The information structure and myopic best response sequence will induce a probability distribution over states and sequences in which players switch. There, a player will have a strict incentive to switch when he is told to switch even if he only expects those before him in the sequence to have switched and has no additional information: this is true because we are averaging across scenarios where the player switches in the myopic best response process. Thus, we have constructed a probability distribution over sequences of recommendations and states for which sequential obedience holds.

Direct information structures (where players are given action recommendations) will in general not be sufficient to smallest equilibrium implement an outcome. Nonetheless, if an outcome satisfies sequential obedience (along with obedience), then we can construct a canonical information structure that implements the outcome in the smallest equilibrium, under a dominance state assumption requiring that in some state, the high action be a dominant action for all players. In this information structure, a sequence is drawn randomly in a way that supports the sequential obedience condition, and in addition a nonnegative integer is drawn randomly according to an almost uniform exponential distribution. Each player’s type is the sum of the integer and his rank in the sequence. Payoffs are re-arranged to ensure that players with first few types have a dominant action to play the high action. One can then argue by induction that if all types of all players up to $\tau$ play the high action, type $\tau + 1$ of any player has a strict incentive to choose the high action: by construction this type will be sure that players with lower ranks than him are playing the high action and hence this is true by sequential obedience.

We provide two applications of our main result. First, we consider the optimization problem of a designer who always prefers players to choose the high action but expects
the worst (hence smallest) equilibrium to be played, and has no instrument other than information design. Due to our characterization result, the problem reduces to a finite linear program. In the case where the game has a potential (Monderer and Shapley (1996)), we identify convexity conditions for the potential and the designer’s objective under which the optimal outcome is one of perfect coordination, where either all players choose the high action or all players choose the low action. Under supermodularity, convexity of the potential is equivalent to requiring that payoffs in the game are not too asymmetric. When these conditions are satisfied, the optimal outcome is easy to characterize. In particular, when the designer cares only about action choices, all players choose the high action conditional on the highest probability event with the property that the resulting expected potential exceeds the expected potential from all playing the low action.

For a second application, we characterize the cheapest way of inducing all players to choose the high action in the smallest (hence unique) equilibrium for a designer who can both design the information structure and offer transfers to the players. Given our characterization result, the problem reduces to deriving the minimum amount of transfers under which the outcome “always play the high action” satisfies sequential obedience.

In the course of studying these applications, we develop a set of simpler characterizations of sequential obedience in potential games that we believe should prove useful in many other applications. An outcome satisfies *coalitional obedience* if no subset of players could increase the potential by simultaneously deviating to playing the low action whenever they were supposed to play the high action. Sequential obedience is equivalent to coalitional obedience in potential games. An outcome satisfies *grand coalitional obedience* if the set of all players could not increase the potential by simultaneously deviating to playing the low action whenever they were supposed to play the high action. A perfectly coordinated outcome satisfies sequential obedience in a convex potential game if and only if it satisfies grand coalitional obedience.

Last, we study full implementability. If an outcome is fully implementable, it satisfies not only sequential obedience which is necessary for smallest equilibrium implementation, but also the reverse version of sequential obedience which is necessary for largest equilibrium implementation. Conversely, we show, under an appropriate extension of
the dominance state assumption, that these necessary conditions are also jointly sufficient for full implementation. We focus on smallest equilibrium implementation rather than full implementation in our analysis for a number of reasons. The arguments for full implementation are mechanical extensions of the results for smallest equilibrium implementation—but more involved to state. Also, smallest equilibrium implementation is the most relevant notion of implementation for our information design application: if an information designer is concerned with inducing the high action in the worst case scenario, smallest equilibrium implementability is the relevant constraint. In fact, we show that if an outcome satisfies sequential obedience and not reverse sequential obedience, then there is another outcome stochastically dominating that outcome that satisfies both sequential obedience and reverse sequential obedience. Thus, if the information designer always prefers higher outcomes, optimal information design subject to full implementability is equivalent to that subject to smallest equilibrium implementability.

1.1. Literature. Our work has its roots in a large literature on the role of higher order beliefs in games. A prominent and early insight in this literature is that certain Nash equilibria of complete information games can be fully implemented via an “infection argument” (Rubinstein (1989) and Carlsson and van Damme (1993)). In particular, consider a two-player two-action game of complete information with two strict Nash equilibria. In a symmetric game, a Nash equilibrium is said to be risk dominant if each player’s action is a best response to a 50/50 conjecture over the actions of the other player. It is possible to construct an incomplete information game where with probability close to 1, payoffs are given by the complete information game, but nonetheless the unique equilibrium of the incomplete information game has the risk dominant action profile of the complete information game played everywhere. Thus we can fully implement the risk dominant outcome by information design with a small perturbation to payoffs. While not always stated like this as an implementation result, this is a lesson that can be drawn from the email game of Rubinstein (1989) and the global games of Carlsson and van Damme (1993). The argument extends to games with asymmetric payoffs for the appropriate definition of risk dominance (Harsanyi and Selten (1988)). Generically, a two-player two-action game with two strict Nash equilibria has exactly one risk dominant equilibrium.
These results showed the sufficiency of risk dominance for full implementation. Kajii and Morris (1997) and Ui (2001) showed necessity: no action profile other than the risk dominant one is fully implementable. Kajii and Morris (1997), Ui (2001), Frankel et al. (2003), and Morris and Ui (2005) generalized these results beyond two-player two-action games. In supermodular games with many players and many actions, Frankel et al. (2003) showed that there are fully implementable outcomes and identified a sufficient condition under which there is a unique fully implementable outcome within global game information structures. The result of Morris and Ui (2005) implies that (a generalized version of) the sufficient condition in Frankel et al. (2003) ensures that no other outcome is fully implementable. Oyama and Takahashi (2020) showed that this sufficient condition is also necessary for generic BAS games: they showed that if the low action profile does not satisfy the sufficient condition identified by Morris and Ui (2005), then some outcome that assigns probability 0 to that action profile is smallest equilibrium implementable, developing a generalization of the classical infection argument.2

These papers addressed the ability to fully implement a particular equilibrium of a complete information game by information design combined with a small perturbation to payoffs. Bergemann and Morris (2019) and Hoshino (2018) report some straightforward applications of these results, emphasizing the information design interpretation. The main result of this paper characterizes what outcomes are smallest equilibrium implementable (and fully implementable) in an incomplete information setting, building on the approach in Oyama and Takahashi (2020). We consider information design without perturbing payoffs, but instead use a dominance state assumption to initiate the infection argument.3

Our first application is to the optimal information design problem where the information designer can choose the information structure but anticipates that the worst

---

2These contributions were not formulated as full implementation results, but rather as characterizing “robustness to incomplete information”. Kajii and Morris (1997) said that a Nash equilibrium of a complete information was robust to incomplete information if every incomplete information game where payoffs are almost always given by that complete information game has a Bayes Nash equilibrium where that Nash equilibrium is almost always played. The infection argument then establishes that the risk dominated equilibrium is not robust.

3Morris and Ui (2019) consider a different incomplete information extension of the literature on robustness to incomplete information described in the previous footnote. They ask: which equilibria of a fixed incomplete information game have the property that nearby outcomes arise in all nearby incomplete information games? Morris and Ui (2019) propose a definition and sufficient conditions that generalize those in Morris and Ui (2005) in general games, addressing subtle issues concerning the correct definition of robustness.
equilibrium will be played ("information design with adversarial equilibrium selection").\textsuperscript{4} Inostroza and Pavan (2020) showed that optimal solutions would satisfy the \textit{perfect coordination property} in a regime change game (an example of a BAS game).\textsuperscript{5} The solution of a symmetric two-player two-action example in Mathevet et al. (2020) also satisfied the perfect coordination property. Contemporaneously with our paper, Li et al. (2019) solved the information design problem considered here in regime change games. Our results provide general conditions under which perfect coordination holds in general BAS games and a general solution which is simple to calculate and interpret. In contrast to these other papers, we show that the perfect coordination property continues to hold even when the games’ payoffs are asymmetric (but not too asymmetric) across players. This contrasts with the partial implementation case, where asymmetric payoffs will lead to optimal solutions not satisfying the perfect coordination property.\textsuperscript{6} In general, there is a large multiplicity of information structures that can be used to implement a particular outcome. Mathevet et al. (2020) and Li et al. (2019) report different and simpler information structures that implement optimal outcomes tailored to specific applications. We provide a canonical information structure that works for all BAS games.

Our second application relates to a literature on inducing effort in teams, which in turn builds on an earlier literature on “divide-and-conquer” strategies in contracting with externalities (Segal (1999) and Winter (2004)). In particular, we show how to generalize a recent contribution of Moriya and Yamashita (2020) on inducing effort in teams in the presence of incomplete information. In doing so, we show a tight connection between this literature and the information design literature. We postpone until Section 4.4 a more detailed discussion of this literature, including a recent paper of Halac et al. (2020) that deals with the complete information case.

Even though our sequential obedience condition has a sequential interpretation, there is no physical sequentiality in our problem. Physical sequentiality is the focus of Doval and Ely (2020), who characterize what outcomes can be partially implementable when

\textsuperscript{4}Early references include Goldstein and Huang (2016), who restrict attention to public information, and Carroll (2016), who considered a bilateral trade game which has binary actions but is not supermodular.

\textsuperscript{5}Inostroza and Pavan (2020) assume that players have exogenous private information, as well as the information provided by a designer. They also work with continuum players and states. None of these differences is important for the role of the perfect coordination property.

\textsuperscript{6}Asymmetric problems give rise to asymmetric solutions in the benchmark model of Arieli and Babichenko (2019), where payoffs are non-strategic and the designer has an intrinsic preference over the correlation of players’ actions. This continues to hold in their strategic extension. We illustrate this asymmetry in our leading example in Section 2.
the actions and payoffs of the players are fixed but the designer can design both the information structure and the extensive form. The version of obedience that arises in their problem is less demanding than the standard static obedience condition whereas our sequential obedience condition is stronger. The obedience condition of Doval and Ely (2020) is weaker than ours both because players assume that later players will follow their recommendations and because they assume that later players’ behavior will respond to their deviations.\footnote{A different sequential version of obedience also appears in Mathevet and Taneva (2020), who study partial implementation when players are constrained to communicate in hierarchies.}

Higher order beliefs matter in our analysis. Mathevet et al. (2020) propose using the (common prior subsets of) universal type space of Mertens and Zamir (1985) to discuss higher order beliefs.\footnote{Sandmann (2020) proposes an alternative representation of the universal type space for information design.} But the different and effective approach in the higher order beliefs literature has been to use belief operators (Monderer and Samet (1989)) on arbitrary common prior type spaces to track higher order beliefs. Morris and Shin (2007) and Morris et al. (2016) introduced generalized belief operators to track those features of higher order beliefs that are relevant for best responses of BAS games. Kajii and Morris (1997) and Oyama and Takahashi (2020) prove results by first establishing properties of higher order beliefs that hold on all common prior type spaces (using belief operators and generalized belief operators, respectively) and then use those properties to establish their results.\footnote{Players’ rank beliefs, i.e., the probability they assign to having a higher expectation than other players, turn out to be crucial. Rank beliefs play an important role in sequential obedience.} However, in this paper, we do not perform the intermediate step of stating results about higher order beliefs on common prior type spaces. This is a pedagogical choice, as we prefer to focus on the obedience interpretation of our necessary and sufficient condition. Nonetheless, it is important to note that common prior properties of higher order beliefs are implicit in our analysis, and could be made explicit with the language of generalized belief operators, following Oyama and Takahashi (2020).

1.2. Setting. There are finitely many players, denoted by \( I = \{1, \ldots, |I|\} \), \(|I| \geq 2\). A state is drawn from a finite set \( \Theta \) according to the probability distribution \( \mu \in \Delta(\Theta) \),\footnote{For a finite or countably infinite set \( X \), we write \( \Delta(X) \) for the set of all probability distributions over \( X \).} where we assume that \( \mu \) has full support: \( \mu(\theta) > 0 \) for all \( \theta \in \Theta \).
Players make binary decisions, \(a_i \in A_i = \{0, 1\}\), simultaneously. We denote \(A = \prod_{i \in I} A_i\) and \(A_{-i} = \prod_{j \neq i} A_j\). Given action profile \(a = (a_i)_{i \in I} \in A\) and state \(\theta \in \Theta\), player \(i \in I\) receives payoff \(u_i(a, \theta)\). Throughout this paper, we assume supermodular payoffs, i.e., for each \(\theta\),

\[
d_i(a_{-i}, \theta) \equiv u_i((1, a_{-i}), \theta) - u_i((0, a_{-i}), \theta)
\]

is weakly increasing in \(a_{-i} \in A_{-i}\). We denote \(\mathbf{0} = (0, \ldots, 0)\) and \(\mathbf{1} = (1, \ldots, 1)\), and write \(\mathbf{0}_{-i}\) and \(\mathbf{1}_{-i}\) for the action profiles of player \(i\)'s opponents such that all players \(j \neq i\) play 0 and 1, respectively. We maintain a dominance state assumption that says that there exists a state where action 1 is a dominant action for all players: there exists \(\bar{\theta} \in \Theta\) such that \(d_i(\mathbf{0}_{-i}, \bar{\theta}) > 0\) for all \(i \in I\). Fixing \(I, A, \) and \(\Theta\), we refer to \((u_i)_{i \in I}\) (or \((d_i)_{i \in I}\)) as the base game.

An information structure is given by a type space \(T = ((T_i)_{i \in I}, \pi)\), in which each \(T_i\) is a countable set of types for player \(i \in I\), and \(\pi \in \Delta(T \times \Theta)\) is a common prior over \(T \times \Theta\), where we write \(T = \prod_{i \in I} T_i\) and \(T_{-i} = \prod_{j \neq i} T_j\). We require an information structure to be consistent with the prior \(\mu\): \(\sum_{t \in T} \pi(t, \theta) = \mu(\theta)\) for each \(\theta \in \Theta\). We also assume that for all \(i \in I\), \(\pi(t_i) \equiv \sum_{t_{-i}, \theta} \pi((t_i, t_{-i}), \theta) > 0\) for all \(t_i \in T_i\).

Together with the base game \((u_i)_{i \in I}\), the information structure \(T\) defines an incomplete information game, which we refer to simply as \(T\). In the incomplete information game \(T\), a strategy for player \(i\) is a mapping \(\sigma_i: T_i \to \Delta(A_i)\). A strategy profile \(\sigma = (\sigma_i)_{i \in I}\) is a (Bayes Nash) equilibrium of the game \(T\) if for all \(i \in I, t_i \in T_i,\) and \(a_i \in A_i\), whenever \(\sigma_i(t_i)(a_i) > 0\), we have

\[
\sum_{t_{-i} \in T_{-i}, \theta \in \Theta} \pi(t_{-i}, \theta|t_i)u_i((a_i, \sigma_{-i}(t_{-i})), \theta) \geq \sum_{t_{-i} \in T_{-i}, \theta \in \Theta} \pi(t_{-i}, \theta|t_i)u_i((a_i', \sigma_{-i}(t_{-i})), \theta)
\]

for all \(a_i' \in A_i\), where \(\pi(t_{-i}, \theta|t_i) = \frac{\pi(t_i, t_{-i}, \theta)}{\pi(t_i)}\), and \(u_i((a_i, \cdot), \theta)\) is extended to \(\prod_{j \neq i} \Delta(A_j)\) in the usual manner. We write \(E(T)\) for the set of equilibria of the game \(T\). Since the game is supermodular, there will always exist a smallest equilibrium, which is in pure strategies (Milgrom and Roberts (1990)). We write \(\sigma(T)\) for that smallest pure strategy equilibrium.

We are interested in induced outcomes, where an outcome is a distribution in \(\Delta(A \times \Theta)\). A pair \((T, \sigma)\) of an information structure and a strategy profile induces outcome \(\nu \in \Delta(A \times \Theta)\):

\[
\nu(a, \theta) = \sum_{t \in T} \pi(t, \theta) \prod_{i \in I} \sigma_i(t_i)(a_i).
\]

An outcome \(\nu\) satisfies consistency if \(\sum_{a \in A} \nu(a, \theta) = \mu(\theta)\) for all \(\theta \in \Theta\).
1.3. **Implementation.** Which outcomes can be implemented by a suitable choice of information structure? The answer will depend on the equilibrium selection rule. We will focus on three different assumptions about equilibrium selection which will give rise to three different notions of implementation.

**Definition 1.** An outcome $\nu \in \Delta(A \times \Theta)$ is partially implementable if there exist an information structure $T$ and an equilibrium $\sigma \in E(T)$ that induce $\nu$.

An outcome $\nu$ satisfies obedience if
\[
\sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu((a_i, a_{-i}), \theta) \left( u_i((a_i, a_{-i}), \theta) - u_i((a_i', a_{-i}), \theta) \right) \geq 0
\]
for all $i \in I$ and $a_i, a_i' \in A_i$.

Bergemann and Morris (2016) showed:

**Proposition 1.** An outcome is partially implementable if and only if it satisfies consistency and obedience.

Bergemann and Morris (2016) called such outcomes Bayes correlated equilibria since they correspond to one natural generalization of correlated equilibrium of Aumann (1974, 1987) to incomplete information games. We write $BCE \subset \Delta(A \times \Theta)$ for the set of Bayes correlated equilibria. Note that $BCE$ is characterized by a finite system of weak linear inequalities and thus is a convex polytope.

This paper characterizes two stronger versions of implementation.

**Definition 2.** Outcome $\nu$ is smallest equilibrium implementable ($S$-implementable) if there exists an information structure $T$ such that $(T, \sigma(T))$ induces $\nu$.

We write $SI \subset \Delta(A \times \Theta)$ for the set of $S$-implementable outcomes. Clearly, $SI \subset BCE$. We characterize $SI$ and its closure $\overline{SI}$ in Section 3. We extend and apply this characterization in Section 4.

**Definition 3.** Outcome $\nu$ is fully implementable if there exists an information structure $T$ such that $(T, \sigma)$ induces $\nu$ for all $\sigma \in E(T)$.$^{11}$

We write $FI \subset \Delta(A \times \Theta)$ for the set of fully implementable outcomes. Clearly, $FI \subset SI$. We report a characterization of $FI$ in Section 5, which is an easy extension of our characterization of $S$-implementation.

$^{11}$ Under supermodularity, full implementation in fact requires $E(T)$ to be a singleton.
2. A Two-State Example

We will use the following example to illustrate ideas throughout the paper. Let us label the action 1 “invest” and the action 0 “not invest”. The payoff to not invest is always 0. There are two states, \( b \) ("bad") and \( g \) ("good"). Each state is equally likely. Payoffs are given by the following tables, where player 1 is the row player and player 2 is the column player:

<table>
<thead>
<tr>
<th></th>
<th>Not Invest</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0, 0</td>
<td>0, −8</td>
</tr>
<tr>
<td>Invest</td>
<td>−7, 0</td>
<td>−4, −5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not Invest</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0, 0</td>
<td>0, 1</td>
</tr>
<tr>
<td>Invest</td>
<td>2, 0</td>
<td>5, 4</td>
</tr>
</tbody>
</table>

Thus both players have a dominant action to invest in the good state and not invest in the bad state. Payoffs are supermodular: a player's payoff to investing is 3 larger if the other player invests. Payoffs are asymmetric: player 1 gets a payoff of 1 larger than player 2 in each scenario.\(^{12}\)

Now consider the outcome represented in the following:

<table>
<thead>
<tr>
<th></th>
<th>Not Invest</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Invest</td>
<td>(\frac{1}{10})</td>
<td>(\frac{2}{5})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not Invest</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Invest</td>
<td>0</td>
<td>(\frac{1}{2})</td>
</tr>
</tbody>
</table>

Evidently, the outcome is consistent. It is also obedient. If player 2 is advised to invest, he assigns probability \(\frac{4}{9}\) to the state being bad and player 1 investing, and he assigns probability \(\frac{5}{9}\) to the state being good and player 1 investing, so his expected payoff to investing is \(\frac{4}{9} \times (-5) + \frac{5}{9} \times 4 = 0\). Thus he is (just) willing to invest. If player 1 is advised to invest, his expected payoff is \(\frac{1}{10} \times (-7) + \frac{2}{5} \times (-4) + \frac{1}{2} \times 5 = \frac{1}{5} > 0\). Thus this outcome is partially implementable (and a Bayes correlated equilibrium). In fact, it is easy to see that this is the best partially implementable outcome for a designer who wants to maximize the expected number of players who invest (which is \(\frac{19}{10}\) under this outcome). It is asymmetric because payoffs are asymmetric. There is slack in the obedience constraint of player 1 so we can sometimes get him to invest even when we cannot get player 2 to invest.

However, if we consider the direct implementation of this outcome, where players are simply advised what action to play, there is a strict equilibrium where both players never invest. If player 1 (the most willing to invest) thinks that player 2 will never invest, his

\(^{12}\)This is an asymmetric variant of an example analyzed in Mathevet et al. (2020).
expected payoff to investing is \(\frac{1}{2}(-7) + \frac{1}{2} \times 2 = -\frac{5}{2} < 0\). No outcome close to this outcome is S-implementable.

Our results will establish that the following outcome is S-implementable (and indeed fully implementable) for all \(0 < \delta \leq \frac{1}{4}\):

\[
\begin{array}{c|cc}
 b & \text{Not} & \text{Invest} \\
\hline
\text{Not} & \frac{1}{4} + \delta & 0 \\
\text{Invest} & 0 & \frac{1}{4} - \delta
\end{array}
\quad (2.2)
\]

and thus that the following outcome is in the closure of the S-implementable set:

\[
\begin{array}{c|cc}
 b & \text{Not} & \text{Invest} \\
\hline
\text{Not} & \frac{1}{4} & 0 \\
\text{Invest} & 0 & \frac{1}{4}
\end{array}
\quad (2.3)
\]

The expected number of players investing is \(\frac{3}{2}\) under the latter outcome, which is in fact the highest expected number of players investing that can be (approximately) S-implemented. More generally, it is the best approximately S-implementable outcome for an information designer who has any convex objective function over the number of players who invest independent of the state. Even though payoffs are asymmetric, it is still optimal to always have perfect coordination, i.e., either both players invest or both players do not invest.

To gain intuition for why this outcome (but no better) is S-implementable, consider the payoffs of the complete information game induced if both players are told to invest. In this case, they would think that the state was good with probability \(\frac{2}{3}\), and payoffs would be

\[
\begin{array}{c|cc}
 & \text{Not} & \text{Invest} \\
\hline
\text{Not} & 0, 0 & 0, -2 \\
\text{Invest} & -1, 0 & 2, 1
\end{array}
\]

The significance of this payoff matrix is that there are two strict Nash equilibria, (Invest, Invest) and (Not Invest, Not Invest), and the (Invest, Invest) equilibrium is just risk dominant as defined by Harsanyi and Selten (1988). We know from thirty year old results in the higher order beliefs literature that it is possible to uniquely implement (Invest, Invest) exactly because it is risk dominant (see Kajii and Morris (1997)) using an “email game information structure” (Rubinstein (1989)). We omit this familiar argument
here, but will later use a modified email information structure both to show that outcome (2.2) is S-implementable and to illustrate the proof of our main result.

We will revisit this example to illustrate sequential obedience in Section 3.1, and show in Section 4 that it has a convex potential (because the asymmetry is not too large), and therefore the optimal outcome satisfies the perfect coordination property if the objective function satisfies a restricted convexity assumption.

3. Smallest Equilibrium Implementation

3.1. Sequential Obedience. We now introduce a strengthening of obedience—which we call *sequential obedience*—that we will show to be necessary and essentially sufficient for S-implementability. Suppose that players’ default action was to play 0 but the information designer recommended a subset of players to play 1, with the designer giving those recommendations sequentially, according to some commonly known distribution on states and sequences of recommendations. When players are advised to play action 1, they will accept the recommendation only if it is a strict best response provided that only players who received the recommendation earlier than them switch.

To describe this formally, let \( \Gamma \) be the set of all sequences of distinct players. For example, if \( I = \{1, 2, 3\} \), then

\[ \Gamma = \{\emptyset, 1, 2, 3, 12, 13, 21, 23, 31, 32, 123, 132, 213, 231, 312, 321\}. \]

For each \( \gamma \in \Gamma \), we denote by \( \bar{a}(\gamma) \in A \) the action profile such that player \( i \) plays action 1 if and only if \( i \) is listed in \( \gamma \). We will call \( \nu_T \in \Delta(\Gamma \times \Theta) \) an *ordered outcome* with the interpretation that \( \nu_T(\gamma, \theta) \) is the probability that the state is \( \theta \), players listed in \( \gamma \) choose action 1 in order \( \gamma \), and players not listed in \( \gamma \) choose action 0. An ordered outcome \( \nu_T \in \Delta(\Gamma \times \Theta) \) *induces* an outcome \( \nu \in \Delta(A \times \Theta) \) in the natural way:

\[ \nu(a, \theta) = \sum_{\gamma: \bar{a}(\gamma) = a} \nu_T(\gamma, \theta). \]

For each \( i \in I \), let \( \Gamma_i \) be the set of all sequences in \( \Gamma \) where player \( i \) is listed. For each \( \gamma \in \Gamma_i \), we denote by \( a_{-i}(\gamma) \in A_{-i} \) the action profile of player \( i \)'s opponents such that player \( j \neq i \) plays action 1 if and only if \( j \) is listed in \( \gamma \) before \( i \) (therefore, player \( j \) plays action 0 if and only if either \( j \) is not listed in \( \gamma \) or \( j \) is listed in \( \gamma \) after \( i \)).

**Definition 4.** An ordered outcome \( \nu_T \in \Delta(\Gamma \times \Theta) \) satisfies *sequential obedience* if

\[ \sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_T(\gamma, \theta)d_i(a_{-i}(\gamma), \theta) > 0 \]  

(3.1)
for all $i \in I$ such that $\nu_\Gamma(\Gamma_i \times \Theta) > 0$. It satisfies weak sequential obedience if the strict inequality in (3.1) is replaced with a weak inequality.

This condition is a restriction on ordered outcomes. However, we want to characterize outcomes (not ordered outcomes) that are implementable, so (abusing notation slightly) we also define sequential obedience as a property of outcomes in the natural way:

**Definition 5.** An outcome $\nu \in \Delta(A \times \Theta)$ satisfies sequential obedience (resp. weak sequential obedience) if there exists an ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ that induces $\nu$ and satisfies sequential obedience (resp. weak sequential obedience).

By definition, the ordered outcome $\nu_\Gamma$ such that $\nu_\Gamma(\emptyset, \theta) = \mu(\theta)$ for all $\theta \in \Theta$ and hence the outcome $\nu$ such that $\nu(0, \theta) = \mu(\theta)$ for all $\theta \in \Theta$ trivially satisfy sequential obedience.

We can illustrate sequential obedience and weak sequential obedience by showing that they are satisfied by outcomes (2.2) and (2.3) in our example in Section 2, respectively.

Consider the ordered outcome $\nu_\Gamma$ given by

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\emptyset$</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>21</td>
<td>$\frac{1}{12}$</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

This satisfies weak sequential obedience (implying that the induced outcome (2.3) also does):

$$\sum_{\gamma \in \Gamma_1, \theta \in \Theta} \nu_\Gamma(\gamma, \theta)d_1(a_1(\gamma), \theta) = \frac{1}{6} \times (-7) + \frac{1}{12} \times (-4) + \frac{1}{3} \times 2 + \frac{1}{6} \times 5 = 0,$$

$$\sum_{\gamma \in \Gamma_2, \theta \in \Theta} \nu_\Gamma(\gamma, \theta)d_2(a_2(\gamma), \theta) = \frac{1}{6} \times (-5) + \frac{1}{12} \times (-8) + \frac{1}{3} \times 4 + \frac{1}{6} \times 1 = 0.$$

Reducing $\nu_\Gamma(12, b)$ or $\nu_\Gamma(21, b)$ by $\delta > 0$ and increasing $\nu_\Gamma(0, b)$ by $\delta$ makes these values strictly positive, and hence the resulting ordered outcome satisfies sequential obedience, as does the induced outcome (2.2).

3.2. **Characterization.** In this section, we show that sequential obedience, along with consistency and obedience, is necessary and essentially sufficient for S-implementability.
Our sufficiency argument will work only for outcomes where all players choose action 1 with positive probability at the dominance state \( \theta \).

**Definition 6.** Outcome \( \nu \) satisfies **grain of dominance** if \( \nu(1, \theta) > 0 \).

We now state our main theorem:

**Theorem 1.**

1. If an outcome is \( S \)-implementable, then it satisfies consistency, obedience and sequential obedience.
2. If an outcome satisfies consistency, obedience, sequential obedience and grain of dominance, then it is \( S \)-implementable.

The proof of Theorem 1 and those of its corollaries below are given in Appendix A.1.

Because of the strict inequalities in the definition of sequential obedience, the set \( SI \) of \( S \)-implementable outcomes is not closed in general. Its closure \( \overline{SI} \), on the other hand, is cleanly characterized by weak sequential obedience. In particular, similar to \( BCE \), \( \overline{SI} \) is a convex polytope.\(^{13}\)

**Corollary 1.** An outcome is contained in \( \overline{SI} \) if and only if it satisfies consistency, obedience, and weak sequential obedience.

Theorem 1 and Corollary 1 require obedience (necessary for partial implementation) as well as (weak) sequential obedience. Note that sequential obedience is stronger than the “upper obedience” requirement that a player want to follow a recommendation to play action 1. If an outcome satisfies sequential obedience, but not the “lower obedience” requirement that a player want to follow a recommendation to play action 0, then, by the construction in the proof of Theorem 1(2), we can find a first-order stochastically dominating outcome in \( SI \) (which satisfies sequential obedience and obedience by Theorem 1(1)).\(^{14}\) Thus we have:

**Corollary 2.**

1. If an outcome \( \nu \) satisfies consistency, sequential obedience, and grain of dominance, then there exists an outcome \( \hat{\nu} \in SI \) that first-order stochastically dominates \( \nu \).

\(^{13}\)The set of ordered outcomes that satisfy weak sequential obedience is characterized by a finite system of weak linear inequalities and thus is a convex polytope. By Corollary 1, \( \overline{SI} \) is the intersection of \( BCE \) and the image of this set under the linear transformation that maps \( \nu_T \in \Delta(\Gamma \times A) \) to \( \nu \in \Delta(A \times \Theta) \) by \( \nu(a, \theta) = \sum_{\gamma: a(\gamma) = a} q_T(\gamma, \theta) \) and hence is a convex polytope.

\(^{14}\)For \( \nu, \hat{\nu} \in \Delta(A \times \Theta) \), we say that \( \hat{\nu} \) first-order stochastically dominates \( \nu \) if for each \( \theta \in \Theta \), \( \hat{\nu}(\cdot, \theta) \) first-order stochastically dominates \( \nu(\cdot, \theta) \): \( \sum_{a' \geq a} \hat{\nu}(a', \theta) \geq \sum_{a' \geq a} \nu(a', \theta) \) for all \( a \in A \).
If an outcome $\nu$ satisfies consistency and weak sequential obedience, then there exists an outcome $\hat{\nu} \in \mathcal{S}^T$ that first-order stochastically dominates $\nu$.

In the remainder of this subsection, we sketch the proof of Theorem 1. First consider necessity (i.e., part (1)). Fix an outcome that is S-implementable. By definition, there must exist an information structure such that the smallest equilibrium induces that outcome. Since the outcome is partially implementable, Proposition 1 implies that it satisfies consistency and obedience.

Now consider a sequence of pure strategy profiles obtained by sequentially taking myopic best responses, starting with the smallest strategy profile. In particular, in each round, pick a player and let all types of that player switch from action 0 to action 1 whenever it is a strict best response to the strategy profile in the previous round. For the choice of player in each round, consider for concreteness a round-robin protocol, where we cycle through players. By supermodularity, the sequence of strategy profiles will be monotone increasing and must converge to the smallest equilibrium, which gives rise to the outcome we have fixed and want to show to satisfy sequential obedience. For each type profile, there will be a set of players who eventually switch to action 1 and there will be a sequence $\gamma$ corresponding to the order in which those players switch. Let us define an ordered outcome by letting the probability of state $\theta$ and sequence $\gamma$ be the probability that $\theta$ is the state and $\gamma$ is the sequence generated by the best response dynamics described above.

By construction, every type who switches to action 1 has a strict incentive to do so, assuming that players before him in the constructed sequence have already switched. In the best response dynamics, a player knows his type and the round. But suppose that he was not told his type or the round, but instead was asked ex ante if he was prepared to always switch to action 1 whenever he would have been told to switch to action 1 under the best response dynamics. We are just averaging across histories where switching to action 1 is a strict best response, so it remains a strict best response even if the player does not know the history. This verifies that the ordered outcome we constructed satisfies sequential obedience: a player knowing that the state and sequence are drawn according to the ordered outcome has a strict incentive to choose action 1 if he expects only players before him in the (unknown to him) realized sequence to play action 1.
Second, consider sufficiency (i.e., part (2)). The proof uses the ordered outcome establishing sequential obedience to construct an information structure where the outcome is induced by the smallest equilibrium. Here we illustrate the construction by showing how to S-implement outcome (2.2) in the example in Section 2. As we noted above, the ordered outcome

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{4} + \delta$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>$\frac{1}{6} - \delta$</td>
<td>$\frac{1}{3}$</td>
</tr>
<tr>
<td>21</td>
<td>$\frac{1}{12}$</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

establishes sequential obedience. Let $\varepsilon > 0$ be sufficiently small that we have

$$\left(\frac{1}{6} - \delta\right) \times (-7) + \frac{1}{12} \times (-4) + \left(\frac{1}{3} - \varepsilon\right) \times 2 + \frac{1}{6} \times 5 > 0,$$

(3.3)

$$\left(\frac{1}{6} - \delta\right) \times (-5) + \frac{1}{12} \times (-8) + \left(\frac{1}{3} - \varepsilon\right) \times 4 + \frac{1}{6} \times 1 > 0.$$  

(3.4)

Then let $\eta > 0$ be much smaller than $\varepsilon$. Now construct information structure $(T, \pi)$ as follows. Let $T_1 = T_2 = \{1, 2, \ldots\} \cup \{\infty\}$, and let $\pi \in \Delta(T \times \Theta)$ be given by

\[
\pi((t_1, t_2), \theta) = \begin{cases} 
\eta(1 - \eta)^m \left(\frac{1}{6} - \delta\right) & \text{if } \theta = b \text{ and } (t_1, t_2) = (m + 1, m + 2) \text{ for some } m \in \mathbb{N}, \\
\eta(1 - \eta)^m \frac{1}{12} & \text{if } \theta = b \text{ and } (t_1, t_2) = (m + 2, m + 1) \text{ for some } m \in \mathbb{N}, \\
\eta(1 - \eta)^m \left(\frac{1}{3} - \varepsilon\right) & \text{if } \theta = g \text{ and } (t_1, t_2) = (m + 1, m + 2) \text{ for some } m \in \mathbb{N}, \\
\eta(1 - \eta)^m \frac{1}{6} & \text{if } \theta = g \text{ and } (t_1, t_2) = (m + 2, m + 1) \text{ for some } m \in \mathbb{N}, \\
\frac{1}{4} + \delta & \text{if } \theta = b \text{ and } (t_1, t_2) = (\infty, \infty), \\
\varepsilon & \text{if } \theta = g \text{ and } (t_1, t_2) = (1, 1), \\
0 & \text{otherwise}; 
\end{cases}
\]

see Table 1. This information structure is generated by the following signal structure:

A nonnegative integer $m$ is drawn according to the distribution $\eta(1 - \eta)^m$. Given the realization of state $\theta$, a sequence $\gamma$ of players is drawn, independently of $m$, according to $\nu_T(\cdot, \theta)$ in (3.2), but with $\nu_T(12, g) - \varepsilon$ in place of $\nu_T(12, g)$. If player $i$ is listed in the sequence $\gamma$, then he receives a signal equal to the sum of $m$ and his ranking in $\gamma$; otherwise, he receives a signal $\infty$. The remaining probability $\varepsilon$ is relocated to $\pi((1, 1), g)$, which will play the role of initiating the infection argument.
We claim that in the smallest equilibrium of this game, both players of types \( t_i < \infty \) will invest. First, each player of type \( t_i = 1 \) assigns probability greater than \( \frac{\varepsilon}{\varepsilon + \eta} \) to the good state, which is close to 1 as \( \eta \ll \varepsilon \), and therefore, invest is a dominant action for this type. Then for \( \tau \geq 2 \), suppose that each player of types \( t_i \leq \tau - 1 \) invests. Given \( \eta \approx 0 \), approximately the payoffs to investing for players 1 and 2 of type \( t_i = \tau \) are then greater than (positive multiplications of)

\[
\frac{1}{12} \times (-4) + \left( \frac{1}{6} - \delta \right) \times (-7) + \frac{1}{6} \times 5 + \left( \frac{1}{3} - \varepsilon \right) \times 2
\]

and

\[
\left( \frac{1}{6} - \delta \right) \times (-5) + \frac{1}{12} \times (-8) + \left( \frac{1}{3} - \varepsilon \right) \times 4 + \frac{1}{6} \times 1,
\]

respectively, which are strictly positive by the conditions (3.3) and (3.4). Therefore, by induction, both players of types \( t_i < \infty \) invest in the smallest equilibrium. Note that players of type \( t_i = \infty \) know that the state is \( b \) and hence do not invest. Thus, the outcome (2.2) is implemented by the smallest (in fact unique) equilibrium of this information structure.
The argument for general BAS games follows identical steps, again using the ordered outcome establishing sequential obedience to construct the type space that S-implements the outcome.

3.3. **Dual Representation.** We conclude this section by reporting a dual representation of the sequential obedience condition. Sequential obedience of an outcome $\nu$ is defined by the existence of an ordered outcome $\nu_T$ inducing $\nu$ that satisfies condition (3.1), or in other words, by the solvability of the system of these equalities and inequalities. A duality theorem thus gives us an equivalent condition in terms of dual variables, as presented in Proposition 2 below. We will use it to prove Proposition 3 in Section 3, where we provide a simpler characterization of sequential obedience when the base game has a potential. It also highlights the important connection with Oyama and Takahashi (2020), as discussed later.

For $\nu \in \Delta(A \times \Theta)$, let $I(\nu) \subset I$ denote the set of “active players” who are recommended to play action 1 with positive probability:

$$I(\nu) = \{ i \in I \mid \nu((1, a_{-i}), \theta) > 0 \text{ for some } a_{-i} \in A_{-i} \text{ and } \theta \in \Theta \}.$$ 

By definition, $\nu(a, \theta) > 0$ only if $S(a) \subset I(\nu)$, where $S(a) = \{ i \in I \mid a_i = 1 \}$.

**Proposition 2.** An outcome $\nu$ satisfies sequential obedience (resp. weak sequential obedience) if and only if for any $(\lambda_i)_{i \in I} \in \mathbb{R}^I_+$ such that $\lambda_i > 0$ for some $i \in I(\nu)$,

$$\sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \max_{\gamma : \bar{a}(\gamma) = a} \sum_{i \in S(a)} \lambda_i d_i(a_{-i}(\gamma), \theta) > (\text{resp.} \geq) 0. \quad (3.5)$$

Thus, sequential obedience requires that for any player weights, the expected weighted sum of payoff changes along the best path be positive. The proof of Proposition 2 is given in Appendix A.2.

For illustration, consider again outcome (2.2) in the example in Section 2, which we denote by $\nu$, where $I(\nu) = I$. For given $(\lambda_i)_{i \in I} \in \mathbb{R}^I_+ \setminus \{0\}$, the left hand side of (3.5) is computed as

$$\left( \frac{1}{4} - \delta \right) \max\{\lambda_1 \times (-7) + \lambda_2 \times (-5), \lambda_2 \times (-8) + \lambda_1 \times (-4) \} + \frac{1}{2} \max\{\lambda_1 \times 2 + \lambda_2 \times 4, \lambda_2 \times 1 + \lambda_1 \times 5 \} = \begin{cases} (\lambda_2 - \lambda_1) \left( \frac{3}{4} + 5\delta \right) + \lambda_1 (12\delta) & \text{if } \lambda_1 \leq \lambda_2, \\ (\lambda_1 - \lambda_2) \left( \frac{3}{2} + 4\delta \right) + \lambda_2 (12\delta) & \text{if } \lambda_1 \geq \lambda_2, \end{cases}$$

18
which is always positive (resp. nonnegative) if $\delta > 0$ (resp. if $\delta = 0$). Thus, Proposition 2 guarantees the existence of some ordered outcome that induces $\nu$ and satisfies sequential obedience (resp. weak sequential obedience) if $\delta > 0$ (resp. if $\delta = 0$).

Oyama and Takahashi (2020) introduced the dual characterization in a related problem. Consider a limiting version of what we study in this paper. In particular, consider a sequence of priors on states where there is probability 1 on a single state in the limit, i.e., $\mu(\theta^*) \to 1$ for some $\theta^* \in \Theta$. In this case, there exists some $\nu \in \Delta(A \times \Theta)$ with $\nu(0, \theta^*) = 0$ that satisfies the weak version of (3.5) if and only if for any $(\lambda_i)_{i \in I} \in \mathbb{R}_+^I$,

$$\max_{a \neq 0} \max_{\gamma(a(\gamma), a^*)} \sum_{i \in S(a)} \lambda_i d_i(a_{-i}(\gamma), \theta^*) \geq 0.$$  

This condition is closely related to the concept of “monotone potential” (Morris and Ui (2005)). In fact, it is equivalent to the condition that there exists no monotone potential for action profile $0$ in the complete information game given by $(d_i(\cdot, \theta^*))_{i \in I}$.\footnote{Formally, a function $v: A \to \mathbb{R}$ is a \textit{monotone potential} for $0$ in the complete information BAS game $(d_i(\cdot, \theta^*))_{i \in I}$ if there exists $(\lambda_i)_{i \in I} \in \mathbb{R}_+^I$ such that $\lambda_i d_i(a_{-i}, \theta^*) \leq v(1, a_{-i}) - v(0, a_{-i})$ for all $i$ and $a_{-i}$, and $v(0) > v(a)$ for all $a \neq 0$.}

Oyama and Takahashi (2020) applied a duality theorem to show that it in turn holds if and only if there exists some ordered outcome $\nu_{\Gamma}$ with $\nu_{\Gamma}(\theta, \theta^*) = 0$ that satisfies weak sequential obedience (3.1). Then they showed that if this condition holds (with strict inequalities by assuming genericity in payoffs), then some outcome $\nu$ with $\nu(0, \theta^*) = 0$ is S-implementable for some sequence of priors $\mu$ with $\mu(\theta^*) \to 1$.\footnote{As reviewed in the Introduction, Morris and Ui (2005) showed, for supermodular games with any finite number of actions, that if the game has a monotone potential that is maximized at an action profile $a^*$ (“monotone potential maximizer”), then $a^*$ is robust to incomplete information (Kajii and Morris (1997)). Oyama and Takahashi (2020) used the result stated in the text to show the converse of this result for generic BAS games.}

In the present paper, in an incomplete information setting, we ask whether a \textit{given} outcome $\nu$ is S-implementable and characterize it by sequential obedience (along with obedience), i.e., the existence of an ordered outcome that induces $\nu$ and satisfies sequential obedience (3.1). Accordingly, its dual condition (3.5) involves the target outcome $\nu$.

4. Applications

Our characterization of S-implementability applies to all BAS games. However, many BAS game applications admit a potential. In this section, we define potential games, derive simpler characterizations of sequential obedience under the existence of a potential and other restrictions, and present our applications.
4.1. Potential Games.

**Definition 7.** The base game \((d_i)_{i \in I}\) is a potential game if there exists \(\Phi : A \times \Theta \to \mathbb{R}\) such that for each \(\theta \in \Theta\),

\[
d_i(a_{-i}, \theta) = \Phi((1, a_{-i}), \theta) - \Phi((0, a_{-i}), \theta)
\]

for each \(i \in I\) and \(a_{-i} \in A_{-i}\).

We will adopt the normalization that \(\Phi(0, \theta) = 0\) for all \(\theta\). We restrict attention to potential games for the remainder of this section.

We will use the following two examples of potential games to illustrate our results. We write \(n(a) = |S(a)|\) for the number of players choosing action 1 in action profile \(a\) and (abusing notation slightly) \(n(a_{-i})\) for the number of players choosing action 1 in action profile \(a_{-i}\).

**Example 1 (Investment Game).** Let \(\Theta = \{1, \ldots, |\Theta|\}\), and

\[
d_i(a_{-i}, \theta) = R(\theta) + h_{n(a_{-i})+1} - c_i,
\]

where \(h_k\) is increasing in \(k\) and \(R(\theta)\) is strictly increasing in \(\theta\). Assume that \(R(|\Theta|)+h_1 > c_i\) for all \(i \in I\), so that the dominance state assumption holds with \(\bar{\theta} = |\Theta|\). We interpret \(d_i(a_{-i}, \theta)\) to be the return to investment (action 1), which is (i) increasing in the state; and (ii) increasing in the proportion of others investing (making the game supermodular).

But there are heterogeneous costs of investment; without loss we assume that

\[
c_1 \leq c_2 \leq \cdots \leq c_{|I|}.
\]

This game has a potential:

\[
\Phi(a, \theta) = R(\theta)n(a) + \sum_{k=1}^{n(a)} h_k - \sum_{i \in S(a)} c_i.
\]

Note that the game \((2.1)\) in Section 2 falls in this class of games with \(R(b) = -7\), \(R(g) = 2\), \(h_1 = 0\), \(h_2 = 3\), \(c_1 = 0\), and \(c_2 = 1\). It has a potential

<table>
<thead>
<tr>
<th>b</th>
<th>Not</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0</td>
<td>-8</td>
</tr>
<tr>
<td>Invest</td>
<td>-7</td>
<td>-12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>g</th>
<th>Not</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Invest</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>
Example 2 (Regime Change Game). Let $\Theta = \{1, \ldots, |\Theta|\}$, and
\[
\begin{align*}
    d_i(a_{-i}, \theta) = \begin{cases} 
    1 - c_i & \text{if } n(a_{-i}) + 1 > |I| - k(\theta), \\
    -c_i & \text{if } n(a_{-i}) + 1 \leq |I| - k(\theta),
    \end{cases}
\end{align*}
\]
where $0 < c_i < 1$, and $k: \Theta \to \mathbb{N}$ is strictly increasing. We assume that $k(1) \geq 1$ and $k(|\Theta|) = |I|$, so that the dominance state assumption holds with $\bar{\theta} = |\Theta|$. The interpretation is that action 0 is to attack the regime while action 1 is to abstain from attacking. The regime collapses if the number of attackers (action 0 players) is greater than or equal to $k(\theta)$, or equivalently, the number of non-attackers (action 1 players) is smaller than $|I| - k(\theta)$. This game has a potential:
\[
\Phi(a, \theta) = \begin{cases} 
    n(a) - (|I| - k(\theta)) - \sum_{i \in S(a)} c_i & \text{if } n(a) > |I| - k(\theta), \\
    -\sum_{i \in S(a)} c_i & \text{if } n(a) \leq |I| - k(\theta).
    \end{cases}
\]

This is a finite-state, finite-player version of the continuous-state, continuum-player regime change game studied of Morris and Shin (2004) and analyzed in this context by Inostroza and Pavan (2020) and Li et al. (2019).

4.2. Simplifying Sequential Obedience. This section provides simpler and more intuitive characterizations of sequential obedience for potential games. Suppose that a subset of players were able to coordinate a deviation where they would always choose action 0, even if action 1 was recommended. We say that an outcome satisfies coalitional obedience if no subset of players would want to deviate in this way. An outcome satisfies grand coalitional obedience if the set of all players would not want to deviate in this way. It is intuitive that such conditions might be relevant for S-implementation, since we are trying to prevent the possibility of joint deviations to a bad equilibrium. These conditions are easier to check than sequential obedience.

In this section, we will show that sequential obedience is equivalent to coalitional obedience in potential games and equivalent to grand coalitional obedience under some extra restrictions. We will use these results in our applications.

4.2.1. Coalitional Obedience. For any outcome $\nu \in \Delta(A \times \Theta)$, define a new potential
\[
\Phi_\nu(a) = \sum_{a' \in A, \theta \in \Theta} \nu(a') \Phi(a \land a', \theta),
\]
where $b = a \land a'$ denotes the action profile such that $b_i = 1$ if and only if $a_i = a'_i = 1$. Imagine that players are told to play action 0 or 1 according to $\nu$. The function $\Phi_\nu$ can be interpreted as a complete information game with a common payoff function $\Phi_\nu$ given.
as follows: Players choose actions $a$, before they receive recommendations $a'$ according to $\nu$. Players who choose action 1 (i.e., those for whom $a_i = 1$) actually play 1 only when recommended so (i.e., $a'_i = 1$), while players who are recommended to play action 0 (i.e., those for whom $a'_i = 0$) are forced to play 0 as “passive players”. Thus, the resulting action profile is $a \land a'$, yielding a payoff $\Phi(a \land a', \theta)$ to every player. Taking the expectation with respect to $\nu$ leads to $\Phi_\nu(a)$.

Recall that, in Section 3, we defined $I(\nu) \subset I$ to be the set of “active players” who are recommended to play action 1 with positive probability under $\nu$. Our coalitional obedience condition requires that the expected potential be maximized when all active players follow the recommendations instead of forming a coalition to deviate to action 0.

**Definition 8.** Outcome $\nu$ satisfies *coalitional obedience* (resp. *weak coalitional obedience*) if

$$\Phi_\nu(1) > (\text{resp. } \geq) \Phi_\nu(a)$$

for all $a \in A$ such that $S(a) \subseteq I(\nu)$.

For illustration, consider again the example in Section 2. For outcome (2.2), which we denote by $\nu$, the average potential $\Phi_\nu$ is given as follows:

<table>
<thead>
<tr>
<th></th>
<th>Not</th>
<th>Invest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>0</td>
<td>$-\frac{3}{2} + 8\delta$</td>
</tr>
<tr>
<td>Invest</td>
<td>$-\frac{3}{4} + 7\delta$</td>
<td>$12\delta$</td>
</tr>
</tbody>
</table>

Thus, outcome $\nu$ satisfies coalitional obedience (resp. weak coalitional obedience) if $\delta > 0$ (resp. if $\delta = 0$).

Our first simplification of sequential obedience is:

**Proposition 3.** In a potential game, an outcome satisfies sequential obedience (resp. weak sequential obedience) if and only if it satisfies coalitional obedience (resp. weak coalitional obedience).

The proof is given in Appendix A.3, where we verify that coalitional obedience is equivalent to the condition given in Proposition 2. The key property is that if the base game is a potential game, the weighted sum of deviation gains across different players is represented by a single function $\Phi_\nu$. 
4.2.2. *Grand Coalitional Obedience and Perfect Coordination.* Grand coalitional obedience is the weaker and simpler requirement that the coalitional obedience condition hold for the set of all players.

**Definition 9.** Outcome $\nu$ satisfies *grand coalitional obedience* (resp. *weak grand coalitional obedience*) if

$$\Phi_\nu(1) > (\text{resp. } \geq) \Phi_\nu(0),$$

or equivalently,

$$\sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \Phi(a, \theta) > (\text{resp. } \geq 0).$$

Thus, grand coalitional obedience is the requirement that the average potential of $\nu$ be greater than that of 0.

Proposition 4 will identify a convexity condition under which sequential obedience reduces to grand coalitional obedience for outcomes where actions are perfectly coordinated. To formalize this, we first define perfect coordination.

**Definition 10.** Outcome $\nu$ satisfies *perfect coordination* if for all $\theta \in \Theta$, $\nu(a, \theta) > 0$ only for $a \in \{0, 1\}$.

As we noted in the introduction, this property played an important role in the regime change game information design analysis of Inostroza and Pavan (2020) and Li et al. (2019), and so will in our applications.

We now introduce the convexity condition for our analysis: that $\Phi(a, \theta)$ is always smaller than a convex combination of $\Phi(0, \theta) = 0$ and $\Phi(1, \theta)$.

**Definition 11.** Potential $\Phi$ satisfies *convexity* if

$$\Phi(a, \theta) \leq \frac{n(a)}{|I|} \Phi(1, \theta)$$

for all $a \in A$ and $\theta \in \Theta$.

This condition requires that payoffs be not too asymmetric across players. To see why, note that if payoffs of the base game are symmetric, so $\Phi(a, \theta) = \Phi(n(a), \theta)$, then supermodularity implies that $\Phi(n + 1, \theta) - \Phi(n, \theta)$ is increasing in $n$ and thus (4.3) is

---

17This condition is thus a strengthening of the requirement that $\arg \max_{a \in A} \Phi(a, \theta) \subset \{0, 1\}$ for all $\theta \in \Theta$. This latter condition is necessary and sufficient for perfect coordination in global game selection in potential games (Frankel et al. (2003)).
satisfied. If payoffs are asymmetric, define a symmetrized potential \( \hat{\Phi} : \{0, \ldots, |I|\} \times \Theta \to \mathbb{R} \) by

\[
\hat{\Phi}(n, \theta) = \frac{1}{\binom{|I|}{n}} \sum_{a : n(a) = n} \Phi(a, \theta).
\]

This represents the average value of the potential \( \Phi(a, \theta) \) across all action profiles where \( n \) players choose action 1. Now a natural measure of the asymmetry of payoffs is

\[
\Delta(a, \theta) = \Phi(a, \theta) - \hat{\Phi}(n(a), \theta).
\]

Here, \( \Delta(a, \theta) \) measures how much higher the value of the potential is for \( a \) relative to the average of actions profiles where the same number of players are choosing action 1. Now supermodularity implies that

\[
M(n, \theta) = \frac{n}{|I|} \Phi(1, \theta) - \hat{\Phi}(n, \theta) \geq 0
\]

for all \( n \) and \( \theta \), where \( M(n, \theta) \) is a measure of the supermodularity of the symmetrized potential. So convexity can be written as the requirement that

\[
\Phi(a, \theta) = \Delta(a, \theta) + \hat{\Phi}(n(a), \theta) \leq \frac{n(a)}{|I|} \Phi(1, \theta)
\]

and so

\[
\Delta(a, \theta) \leq M(n(a), \theta)
\]

for any \( a \in A \) and \( \theta \in \Theta \).

We can illustrate convexity and its “not too much heterogeneity” interpretation with our examples:

**Example 3** (Investment Game). In the game as defined in Example 1, convexity holds if and only if

\[
\frac{1}{\ell} \sum_{k=1}^{\ell} (h_k - c_k) \leq \frac{1}{|I|} \sum_{k=1}^{|I|} (h_k - c_k)
\]

for any \( \ell = 1, \ldots, |I| - 1 \). This condition automatically holds if costs are symmetric and amounts to the assumption that costs are not too asymmetric. In particular, a sufficient condition for convexity is that:

\[
h_k - c_k \leq h_{k+1} - c_{k+1}
\]

for any \( k = 1, \ldots, |I| - 1 \), where \( h_k \) is increasing by supermodularity.

**Example 4** (Regime Change Game). In the game as defined in Example 2, convexity holds if and only if \( c_1 = \cdots = c_{|I|} \).
Now we have:

**Proposition 4.** Suppose that the potential satisfies convexity. Then a perfectly coordinated outcome satisfies sequential obedience (resp. weak sequential obedience) if and only if it satisfies grand coalitional obedience (resp. weak grand coalitional obedience).

The proof is given in Appendix A.4.

4.3. **Application 1: Information Design with Adversarial Equilibrium Selection.** Characterizations of implementability are key ingredients in optimal information design problems. Suppose that an information designer receives $V(a, \theta)$ if players choose $a \in A$ in state $\theta \in \Theta$. We maintain the *monotonicity* assumption on $V$: for each $\theta \in \Theta$, $V(a, \theta)$ is weakly increasing in $a$.

We will adopt the normalization that $V(0, \theta) = 0$. For simplicity, we assume that $V(1, \theta) > 0$ for all $\theta \in \Theta$. We are interested in the optimal information design problem with adversarial equilibrium selection, where the designer wants to obtain the best possible payoffs even if players will play her worst equilibrium, which, by the monotonicity of $V$ in $a$, is the smallest equilibrium $\sigma = \sigma(T)$. Thus her problem is:

$$\sup_T \min_{\sigma \in E(T)} \sum_{t \in T, \theta \in \Theta} \pi(t, \theta)V(\sigma(t), \theta) = \sup_T \sum_{t \in T, \theta \in \Theta} \pi(t, \theta)V(\sigma(t), \theta),$$

where $V(\cdot, \theta)$ is extended to $\Pi_{i \in I} \Delta(A_i)$ in the usual manner. Under our definition of S-implementable outcomes, this is equivalent to

$$\sup_{\nu \in \overline{\Delta}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta) = \max_{\nu \in \overline{\Delta}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta).$$

Now an *optimal outcome* of the adversarial information design problem is any element $\nu$ of $\overline{\Delta}$ that maximizes $\sum_{a, \theta} \nu(a, \theta)V(a, \theta)$. By Corollary 1, we immediately have the following:

**Corollary 3.** An outcome is an optimal outcome of the adversarial information design problem if and only if it is an optimal solution to the problem

$$\max_{\nu \in \Delta(A \times \Theta)} \sum_{a, \theta} \nu(a, \theta)V(a, \theta)$$

subject to consistency, obedience, and weak sequential obedience.

By Corollary 2, an optimal outcome of the adversarial information design problem can be obtained by a maximal (with respect to first-order stochastic dominance) optimal
solution to the relaxed problem $\max_{\nu \in \Delta(A \times \Theta)} \sum_{a, \theta} \nu(a, \theta) V(a, \theta)$ subject to consistency and weak sequential obedience (without obedience imposed).

Our characterization of optimal outcomes requires one additional assumption on the designer’s objective:

**Definition 12.** Designer’s objective $V$ satisfies *restricted convexity* with respect to potential $\Phi$ if

$$V(a, \theta) \leq \frac{n(a)}{|I|} V(1, \theta)$$

whenever $\Phi(a, \theta) > \Phi(1, \theta)$.

Convexity of $V$, $V(a, \theta) \leq \frac{n(a)}{|I|} V(1, \theta)$ for all $a$ and $\theta$, is obviously a sufficient condition for restricted convexity, irrespective of $\Phi$. As discussed above when discussing the convexity of $\Phi$, we can say more about convexity when $V$ is supermodular. In this case, convexity of $V$ is equivalent to the assumption that the designer does not distinguish among players too much; and convexity holds automatically if players are treated identically. Thus, for example, convexity holds if $V(a, \theta) = \left(\frac{n(a)}{|I|}\right)\alpha$ with $\alpha \geq 1$; in particular, when the designer wants to maximize the expected fraction of players who play action 1 ($\alpha = 1$), or the probability that all players play 1 ($\alpha \to \infty$).

An important setting where convexity fails but restricted convexity holds is in regime change games where the designer’s objective is to maximize the probability of regime change.\(^{18}\)

**Example 5** (Regime Change Game). In the game as defined in Example 2, $\Phi(a, \theta) > \Phi(1, \theta)$ holds only when $n(a) \leq |I| - k(\theta)$ (i.e., when the regime collapses). Thus, $V$ satisfies restricted convexity with respect to $\Phi$, for example, if

$$V(a, \theta) = \begin{cases} 1 & \text{if } n(a) > |I| - k(\theta), \\ 0 & \text{if } n(a) \leq |I| - k(\theta). \end{cases}$$

Assume that the potential $\Phi$ satisfies convexity and the objective $V$ satisfies restricted convexity with respect to $\Phi$. Theorem 2 will establish that there is an optimal outcome that satisfies perfect coordination. Once we know that the solution satisfies perfect coordination, due to Proposition 4 it is easy to characterize such an optimal outcome, and we first do so.

\(^{18}\)This is the objective of the information designer in the regime change applications of Inostroza and Pavan (2020) and Li et al. (2019). Inostroza and Pavan (2020) also study a richer class of regime change games (in Section 3) and identify a restriction on the designer’s objective (condition PC) that allows their analysis to go through.
Consider the maximization problem with respect to perfectly coordinated outcomes subject to consistency and weak grand coalitional obedience:

$$\max_{\nu(1, \theta) \in \Theta} \sum_{\theta \in \Theta} \nu(1, \theta)V(1, \theta)$$  \hspace{1cm} (4.7a)

subject to

$$\sum_{\theta \in \Theta} \nu(1, \theta)\Phi(1, \theta) \geq 0,$$  \hspace{1cm} (4.7b)

$$0 \leq \nu(1, \theta) \leq \mu(\theta) \quad (\theta \in \Theta).$$  \hspace{1cm} (4.7c)

To solve this problem, we relabel the states as $\Theta = \{1, \ldots, |\Theta|\}$ in such a way that $\Phi(1, \theta)/V(1, \theta)$ is increasing in $\theta$:

$$\frac{\Phi(1, 1)}{V(1, 1)} \leq \cdots \leq \frac{\Phi(1, |\Theta|)}{V(1, |\Theta|)}.$$  \hspace{1cm} (4.8)

By the dominance state assumption, $\Phi(1, \theta) > 0$. Then define

$$\Psi(\theta) = \sum_{\theta' \geq \theta} \mu(\theta')\Phi(1, \theta').$$

If $\Psi(1) \geq 0$, then the outcome “all play 1” is an optimal solution. In the following, we assume that $\Psi(1) < 0$. Let $\theta^* \in \Theta$ be a unique state such that $\Psi(\theta) \geq 0$ if and only if $\theta \geq \theta^*$. Note that $\theta^* > 1$ and $\Phi(1, \theta^* - 1) < 0$. Let

$$p^* = \frac{\Psi(\theta^*)}{-\Phi(1, \theta^* - 1)}.$$  \hspace{1cm} (4.9)

By construction, $0 \leq p^* < \mu(\theta^* - 1)$; indeed, we have $p^* \geq 0$ since $\Psi(\theta^*) \geq 0$, and $p^* - \mu(\theta^* - 1) = \Psi(\theta^* - 1)/(-\Phi(1, \theta^* - 1)) < 0$ since $\Psi(\theta^* - 1) < 0$.

Now define the perfectly coordinated outcome $\nu^*$ by

$$\nu^*(a, \theta) = \begin{cases} 
\mu(\theta) & \text{if } a = 1 \text{ and } \theta \geq \theta^*, \\
p^* & \text{if } a = 1 \text{ and } \theta = \theta^* - 1, \\
\mu(\theta) - p^* & \text{if } a = 0 \text{ and } \theta = \theta^* - 1, \\
\mu(\theta) & \text{if } a = 0 \text{ and } \theta < \theta^* - 1, \\
0 & \text{otherwise}, 
\end{cases}$$  \hspace{1cm} (4.10)

which clearly satisfies consistency (4.7c). This satisfies the weak grand coalitional obedience condition (4.7b) with equality:

$$\sum_{a \in A, \theta \in \Theta} \nu^*(a, \theta)\Phi(a, \theta) = \Psi(\theta^*) + p^*\Phi(1, \theta^* - 1) = 0.$$  \hspace{1cm} (4.11)

It also satisfies lower obedience: for all $i \in I$,

$$\sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu^*((0, a_{-i}), \theta)d_i(a_{-i}, \theta) = \sum_{\theta \leq \theta^* - 1} \nu^*(0, \theta)\Phi((1, 0_{-i}), \theta) \leq 0;$$
since by the convexity of $\Phi$, $\Phi((1, 0_{-i}), \theta) \leq \frac{1}{|I|} \Phi(1, \theta) < 0$ for all $\theta \leq \theta^* - 1$. Thus, $\nu^* \in \overline{S_I}$ by Proposition 4 and Corollary 1. Theorem 2 shows that $\nu^*$ is an optimal solution to the problem (4.7) and that it is an optimal outcome of the adversarial information design problem.

**Theorem 2.** Suppose that $\Phi$ satisfies convexity and $V$ satisfies restricted convexity with respect to $\Phi$. Then there exists an optimal outcome of the adversarial information design problem that satisfies perfect coordination. In particular, the outcome $\nu^*$ defined in (4.10) is an optimal outcome.

The proof is given in Appendix A.5.

We can illustrate the result with the two-player example in Section 2. Suppose that the designer wants to maximize the expected number of players who invest, i.e., $V(a, \theta) = n(a)$, so that restricted convexity is satisfied. With the potential $\Phi$ given in Example 1, we have $\Psi(b) = -3$ and $\Psi(g) = 3$, and hence $\theta^* - 1 = b$ and $\theta^* = g$. With $p^* = \frac{1}{4}$, the optimal outcome $\nu^*$ is thus given in (2.3) in Section 2.

It is worth highlighting assumptions that we have not made in this application. The only restrictions on the designer’s payoff are monotonicity and restricted convexity, so in particular, it may be state-dependent. Also, we did not make any monotonicity assumption about how payoffs depend on the state. However, if we consider the often-studied case where $V$ is independent of $\theta$, the order (4.8) on the states is determined only by the base game through $\Phi(1, \theta)$. The characterization then becomes yet simpler in a continuous-state version of the problem where $\Theta = \mathbb{R}$ and $d_i(a_{-i}, \theta)$ (hence $\Phi(1, \theta)$) is increasing in $\theta$. In this case, the threshold $\theta^*$ will be the unique solution to

$$\int_{\theta \geq \theta^*} \Phi(1, \theta)f(\theta)d\theta = 0,$$

where $f$ denotes the density of $\theta$.

The general construction in the proof of Theorem 1 provides an information structure that implements the optimal outcome identified in Theorem 2. However, the general construction (covering non-potential games) depends on fine details of the game. Under the maintained assumption in this section that the game has a potential, it is possible to provide simpler implementations that do not depend on the fine details of the game. In a continuous-state model, Frankel et al. (2003) showed that one can fully implement
a potential maximizing action using canonical global game information structures (i.e., having players observe the state with small conditionally independent additive noise).

4.4. Application 2: Adding Transfers. In the previous application, we proved general results about the pure information design problem using our simplified characterizations of sequential obedience in Section 4.2. In this application, we also use those characterizations of sequential obedience, but now ask what can be achieved if the designer can use transfers as well as information design. Specifically, we consider the problem of fully implementing effort in team problems in an incomplete information version of Winter (2004).

There is a team $I$ of agents who are engaged in a joint project. Each agent decides whether to exert effort (action 1) or not (action 0), where the effort cost is $c > 0$, common across agents. The probability of success of the project depends on the number of agents who exert effort as well as the state of the world, unknown to the agents, which is drawn from a finite set $\Theta$ according to a full support distribution $\mu$. The project’s technology is given by the function $p: \{0, \ldots, |I|\} \times \Theta \rightarrow [0, 1]$, where $p(n, \theta)$ is the success probability when $n$ agents exert effort at state $\theta$. We assume that $p(n, \theta)$ is nondecreasing in $n$ for each $\theta \in \Theta$, and $p(|I|, \theta) > p(0, \theta)$ for some $\theta \in \Theta$ (so that $\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta)) > 0$).

For $n = 1, \ldots, |I|$, write

$$\Delta p(n, \theta) = p(n, \theta) - p(n-1, \theta).$$

We also assume increasing returns to scale (IRS) on $p$, i.e., we assume that $\Delta p(n, \theta)$ is nondecreasing in $n$ for each $\theta \in \Theta$.

The principal chooses an information structure (as in the previous section) and a bonus payment scheme. We assume that the actions of the agents are unobservable and the state realization is unverifiable, so that the bonus payment to each agent can depend only on the success of the project. If the bonus payment to agent $i$ is $b_i > 0$, this agent’s payoff is thus given by $p(n(a_i - 1) + 1, \theta)b_i - c$ for $a_i = 1$ and $p(n(a_i), \theta)b_i$ for $a_i = 0$. By normalization, we let the payoff difference function be given by

$$d_i(a_{-i}, \theta; b_i) = \Delta p(n(a_{-i}) + 1, \theta) - \frac{c}{b_i}.$$ 

By the assumption of IRS, $d_i$ is nondecreasing in $a_{-i}$.

The objective of the principal is to find a bonus scheme $b = (b_i)_{i \in I}$ and an information structure that minimize the total payment while inducing all types of all agents to exert
effort in the smallest, hence unique, equilibrium. Thus, the problem becomes:

$$\inf_{b, \nu \in \text{SI}(b)} \sum_{i \in I} b_i,$$

where $\nu \in \Delta(A \times \Theta)$ is the “always play 1” outcome, i.e., the outcome such that $\nu(1, \theta) = \mu(\theta)$ for all $\theta \in \Theta$, and $\text{SI}(b) \subseteq \Delta(A \times \Theta)$ is the set of $S$-implementable outcomes under the bonus scheme $b$. We say that a bonus scheme $b^* = (b_i^*)_{i \in I}$ is optimal if $\sum_{i \in I} b_i^*$ is equal to this infimum and $\nu \in \text{SI}(b^* + \varepsilon)$ for every $\varepsilon > 0$, where $b^* + \varepsilon = (b_i^* + \varepsilon)_{i \in I}$.

Let $\bar{\theta} \in \Theta$ be a state such that $\Delta p(1, \bar{\theta}) \geq \Delta p(1, \theta)$ for all $\theta \in \Theta$. We impose the following assumption:

$$\Delta p(1, \bar{\theta}) \geq \sum_{\theta \in \Theta} \mu(\theta) \frac{p(|I|, \theta) - p(0, \theta)}{|I|}. \quad (4.12)$$

This corresponds to the dominance state assumption in the main analysis. It says that the marginal productivity by a single agent’s effort at $\bar{\theta}$ (left hand side) is large enough that it exceeds the expected average productivity (right hand side). Under this condition, we derive the optimal bonus scheme.

By Theorem 1(1), sequential obedience of $\nu$ under $(d_i(\cdot; b_i))_{i \in I}$ is a necessary condition for $\nu \in \text{SI}(b)$, which will give us a condition on payoffs, hence bonuses. But the base game given $b = (b_i)_{i \in I}$ is a potential game with a potential

$$\Phi(a, \theta; b) = p(n(a), \theta) - p(0, \theta) - \sum_{i \in S(a)} \frac{c}{b_i}.$$

Therefore, by Proposition 3, sequential obedience reduces to the simpler condition of coalitional obedience. For the outcome $\nu$, coalitional obedience is equivalent to the condition that the action profile $1$ is a unique maximizer of the average potential: $\sum_{\theta \in \Theta} \mu(\theta) \Phi(1, \theta; b) > \sum_{\theta \in \Theta} \mu(\theta) \Phi(a, \theta; b)$ for all $a \neq 1$. We consider the minimization problem under the relaxed constraint of weak grand coalitional obedience $\sum_{\theta \in \Theta} \mu(\theta) \Phi(1, \theta; b) \geq \sum_{\theta \in \Theta} \mu(\theta) \Phi(0, \theta; b) = 0$:

$$\min_b \sum_{i \in I} b_i$$

subject to

$$\sum_{i \in I} \frac{c}{b_i} \leq \sum_{\theta \in \Theta} \mu(\theta) (p(|I|, \theta) - p(0, \theta)).$$

Since the left hand side of the constraint, which is binding, is (continuous and) strictly quasi-convex in $b$ (by the strict convexity of $x \mapsto \frac{1}{x}$), an optimal solution to this relaxed
problem is unique. It is readily verified to be \( b^* = (\beta^*, \ldots, \beta^*) \) with

\[
\beta^* = \frac{|I|c}{\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta))}.
\] (4.13)

This will indeed be a (unique) optimal bonus scheme if \( \bar{\nu} \in SI(b^* + \varepsilon) \) for every \( \varepsilon > 0 \). Under \( b^* + \varepsilon \), the potential is now

\[
\Phi(a, \theta; b^* + \varepsilon) = p(n(a), \theta) - p(0, \theta) - n(a) \frac{c}{\beta^* + \varepsilon},
\]

which satisfies convexity by IRS. Therefore, by Proposition 4, sequential obedience is equivalent to grand coalitional obedience \( \sum_{\theta \in \Theta} \mu(\theta)\Phi(1, \theta; b^* + \varepsilon) > \sum_{\theta \in \Theta} \mu(\theta)\Phi(0, \theta; b^* + \varepsilon) = 0 \), which is clearly satisfied for any \( \varepsilon > 0 \). Lower obedience is trivially satisfied for the outcome \( \bar{\nu} \). Finally, by the assumption (4.12), we have \( \beta^* \geq \frac{c}{\Delta p(1, \theta)} \), and therefore,

\[
d_i(0-I, \bar{\theta}; b^*_i + \varepsilon) = \Delta p(1, \bar{\theta}) - \frac{c}{\beta^* + \varepsilon} > 0,
\]

so that the dominance state assumption is satisfied for any \( \varepsilon > 0 \). Hence, from Theorem 1(2), it follows that \( \bar{\nu} \in SI(b^* + \varepsilon) \) for any \( \varepsilon > 0 \). Thus, we have:

**Proposition 5.** The unique optimal bonus scheme is given by \( b^* = (\beta^*, \ldots, \beta^*) \), where \( \beta^* \) is as defined in (4.13).

Let us close this section by briefly discussing related studies. The original model of Winter (2004) has no state uncertainty (or \(|\Theta| = 1\)). Winter (2004) shows that, even with symmetric effort costs, an optimal bonus scheme must be discriminatory. Specifically, it is given by \( \left( \frac{c}{\Delta p(1)}, \ldots, \frac{c}{\Delta p(|I|)} \right) \) (modulo permutation). Moriya and Yamashita (2020) introduce state uncertainty to Winter’s (2004) model and study the joint design of bonus payments and information allocation with two agents and two states. They derive the optimal bonus scheme restricting to symmetric schemes \( b_1 = b_2 \). Our analysis extends theirs to any (finite) numbers of agents and states and shows that, under the dominance state assumption (i.e., assumption (4.12)), a symmetric bonus scheme is indeed optimal among asymmetric schemes, and asymmetric ones are strictly suboptimal. A recent paper of Halac et al. (2020) also considers the joint design of payments and information in Winter’s (2004) model (with many agents and no state uncertainty, but with possible asymmetries in production technology and costs), but where the payments may vary across types within the information structure (in our terminology, \( b_i \) is a function on \( T_i \)). They derive the optimal value of the total payment and the optimal “ranking scheme”, which is analogous to the construction in our Theorem 1(2) (and more closely to that
in Oyama and Takahashi (2020)). In particular, they show that in the limit as the infimum of the total payment is approached, the bonus distribution collapses to a point mass on a certain profile of bonuses. By our method applied to their setting, in fact, one can show that the limit optimal bonus profile is characterized as the solution to the payment minimization problem subject to the condition that the action profile \( \mathbf{1} \) (“all agents working”) is a maximizer of the potential of the corresponding limit complete information game. With sufficiently asymmetric technology and costs, in general, the potential will not be convex, and thus the grand coalitional obedience condition, that \( \mathbf{1} \) has a larger potential value than \( \mathbf{0} \) (“all agents shirking”), will not be the only binding constraint. The comparative statics conducted in Halac et al. (2020) with respect to costs (under symmetric technology) may be understood from this viewpoint as determining the binding constraints.

5. Full Implementation

We focussed on S-implementation, rather than full implementation, because it is the most relevant notion for applications and is simpler to state. Indeed, our argument easily extends to full implementation. We show that both sequential obedience and its reversed version are necessary and jointly sufficient for full implementation. We also show that, under the monotonicity assumption on \( V \), the adversarial information design problem as studied in Section 4.3 is equivalent to the information design problem with full implementation.

To proceed, we add a symmetric dominance state assumption that there exists \( \bar{\theta} \in \Theta \) such that \( d_i(1_{-i}, \bar{\theta}) < 0 \) for all \( i \in I \). We now allow an alternative interpretation of an ordered outcome as describing switches from action 1 to action 0. Thus, for a sequence \( \gamma^0 \in \Gamma \), write \( \bar{a}^0(\gamma^0) = 1 - a(\gamma^0) \in A \) for the action profile such that player \( i \) plays action 0 if and only if \( i \) is listed in \( \gamma^0 \) and \( a^0_{-i}(\gamma^0) \in A_{-i} \) for the action profile such that only players before \( i \) in \( \gamma^0 \) play action 0. Thus an ordered outcome \( \nu^0_1 \in \Delta(\Gamma \times \Theta) \) reverse induces \( \nu \in \Delta(A \times \Theta) \) if

\[
\nu(a, \theta) = \sum_{\gamma^0: a^0(\gamma^0) = a} \nu^0_1(\gamma^0, \theta).
\]

**Definition 13.** An ordered outcome \( \nu^0_1 \in \Delta(\Gamma \times \Theta) \) satisfies reverse sequential obedience if

\[
\sum_{\gamma^0 \in \Gamma, \theta \in \Theta} \nu^0_1(\gamma^0, \theta) d_i(a^0_{-i}(\gamma^0), \theta) < 0 \quad (5.1)
\]
for all $i \in I$ such that $\nu_1^0(\Gamma_i \times \Theta) > 0$. An outcome $\nu \in \Delta(A \times \Theta)$ satisfies reverse sequential obedience if there exists an ordered outcome $\nu_1^0 \in \Delta(\Gamma \times \Theta)$ that reverse induces $\nu$ and satisfies reverse sequential obedience.

**Definition 14.** Outcome $\nu$ satisfies two-sided grain of dominance if $\nu(1, \theta) > 0$ and $\nu(0, \theta) > 0$.

**Theorem 3.** (1) If an outcome is fully implementable, then it satisfies consistency, sequential obedience, and reverse sequential obedience.

(2) If an outcome satisfies consistency, sequential obedience, reverse sequential obedience, and two-sided grain of dominance, then it is fully implementable.

Necessity (i.e., part (1)) follows immediately by applying Theorem 1(1) in both directions. The proof for sufficiency (i.e., part (2)), given in Supplemental Appendix B.1, is a simple adaption of the proof of Theorem 1(2). It proceeds with two ordered outcomes satisfying sequential obedience and reverse sequential obedience, respectively. Then we construct an information structure analogous to that used in Theorem 1(2), where an integer was drawn almost uniformly on the integers and a player observed a signal equal to that integer plus his rank in the sequence drawn from the ordered outcome establishing sequential obedience. But now two sequences are drawn independently from the two ordered outcomes (conditional on the recommended action profile and the state). A player’s type will consist of an integer signal and an action recommendation, where the recommended action indicates which of the two sequences generates the integer signal. Then, an induction argument analogous to that for Theorem 1(2) shows that there is a unique equilibrium (in fact, unique rationalizable strategy profile), which induces the target outcome.

Clearly, full implementability is stronger than S-implementability. Yet, we show that maximal S-implementable outcomes (with respect to first-order stochastic dominance) must indeed be fully implementable.

**Proposition 6.** For any $\nu \in \overline{SI}$, there exists $\hat{\nu} \in \overline{FI}$ that first-order stochastically dominates $\nu$.

---

19For the game in Section 2, for example, if we reverse the order on actions to be “Invest < Not Invest”, outcome (2.2) with $-\frac{1}{4} \leq \delta < 0$ satisfies sequential obedience but not reverse sequential obedience and hence is S-implementable but not fully implementable.
The proof is given in Supplemental Appendix B.2.

Consider, as in Section 4.3, the optimal information design problem but with full implementation:

$$\sup_{\nu \in \mathcal{F}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta) = \max_{\nu \in \mathcal{F}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta).$$

By Proposition 6, under the monotonicity assumption on $V$, solving this problem amounts to solving the problem with S-implementation: we have

$$\max_{\nu \in \mathcal{F}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta) = \max_{\nu \in \mathcal{S}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta)V(a, \theta),$$

and by Corollary 2, an optimal outcome of the information design problem with full implementation can be obtained by a maximal optimal solution to the problem

$$\max_{\nu \in \Delta(A \times \Theta)} \sum_{a, \theta} \nu(a, \theta)V(a, \theta) \text{ subject to consistency and weak sequential obedience.}$$

**APPENDIX**

A.1. **Proof of Theorem 1.**

A.1.1. **Proof of Theorem 1(1).** In this proof, since the strategies to appear are all pure, by abusing notation we let $\sigma_i(t_i)$ represent a pure action (an element of $A_i$), rather than a mixed action (an element of $\Delta(A_i)$).

Let $\nu \in \Delta(A \times \Theta)$ be S-implementable, and let $(T, \pi)$ be a type space whose smallest equilibrium $\sigma$ induces $\nu$. By Proposition 1, $\nu$ satisfies consistency and obedience.

Consider the sequence of pure strategy profiles $\{\sigma^n\}$ obtained by sequential best response starting with the smallest strategy profile: let $\sigma^n_i(t_i) = 0$ for all $i \in I$ and $t_i \in T_i$, and for round $n = 1, 2, \ldots$, all types of player $n \pmod{|I|}$ switch from action 0 to action 1 if it is a strict best response to $\sigma^n_{i-1}(t_i)$. Thus,

$$\sigma^n_i(t_i) = \begin{cases} 1 & \text{if } i \equiv n \pmod{|I|}, \\ \sigma^n_{i-1}(t_i) & \text{otherwise.} \end{cases}$$

By supermodularity, for each $i$ and $t_i$, the sequence $\{\sigma^n_i(t_i)\}$ (of pure actions 0 and 1) is monotone increasing and converges to $\sigma_i(t_i)$. Let $n_i(t_i) = n$ if $\sigma^n_{i-1}(t_i) = 0$ and $\sigma^n_i(t_i) = 1$ (and hence $\sigma_i(t_i) = 1$); let $n_i(t_i) = \infty$ if $\sigma^n_i(t_i) = 0$ for all $n$ (and hence $\sigma_i(t_i) = 0$). Write $n(t) = (n_1(t_1), \ldots, n_\ell(t_\ell))$. For $\gamma = (i_1, \ldots, i_k) \in \Gamma$, let $T(\gamma)$ denote the set of type profiles $t$ such that $n(t)$ is ordered according to $\gamma$: i.e., those type profiles $t$ such that $n_i(t_i) = \infty$ for all $i \notin \{i_1, \ldots, i_k\}$, and $n_{i_\ell}(t_{i\ell}) < n_{i_m}(t_{i_m}) < \infty$ if and only if $\ell < m$. 

34
Now, define \( \nu_T \in \Delta(\Gamma \times \Theta) \) by
\[
\nu_T(\gamma, \theta) = \sum_{t \in T(\gamma)} \pi(t, \theta)
\]
for each \((\gamma, \theta) \in \Gamma \times \Theta\). Observe that \( \nu_T \) induces \( \nu \): indeed, for each \((\gamma, \theta) \in \Gamma \times \Theta\), we have
\[
\sum_{\gamma: \bar{a}(\gamma) = a} \nu_T(\gamma, \theta) = \sum_{\gamma: \bar{a}(\gamma) = a} \sum_{t \in T(\gamma)} \pi(t, \theta) = \sum_{t: n_i(t_i) < \infty} \pi(t, \theta) = \nu(a, \theta).
\]
To show sequential obedience, fix any \( i \in I \) with \( \nu_T(\Gamma_i \times \Theta) > 0 \). Note that for all \( t_i \in T_i \) with \( n_i(t_i) < \infty \), we have
\[
\sum_{t_i: n_i(t_i) < \infty} \nu_T(t_i, t_{-i}, \theta) d_i \left( \sigma_{n_i(t_i)}^{-1}(t_{-i}), \theta \right) > 0.
\]
By adding up the inequality over all such \( t_i \), we have
\[
0 < \sum_{t_i: n_i(t_i) < \infty, t_{-i} \in T_{-i}, \theta \in \Theta} \sum \pi((t_i, t_{-i}), \theta) d_i \left( \sigma_{n_i(t_i)}^{-1}(t_{-i}), \theta \right)
\]
\[
= \sum_{\gamma \in \Gamma_i, t \in T(\gamma)} \sum_{\theta \in \Theta} \pi(t, \theta) d_i (a_{-i}(\gamma), \theta)
\]
\[
= \sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_T(\gamma, \theta) d_i (a_{-i}(\gamma), \theta).
\]
Thus, \( \nu \) satisfies sequential obedience.

A.1.2. Proof of Theorem 1(\(2)\). Let \( \nu \in \Delta(A \times \Theta) \) satisfy consistency, obedience, sequential obedience, and grain of dominance, and let \( \nu_T \in \Delta(\Gamma \times \Theta) \) be an ordered outcome establishing sequential obedience. Since \( \nu(1, \bar{\theta}) > 0 \) by grain of dominance, there exists \( \bar{\gamma} \in \Gamma \) containing all players with \( \nu_T(\bar{\gamma}, \bar{\theta}) > 0 \). For \( \varepsilon > 0 \) with \( \varepsilon < \nu_T(\bar{\gamma}, \bar{\theta}) \), let
\[
\hat{\nu}_T(\gamma, \theta) = \begin{cases} 
\nu_T(\gamma, \theta) - \varepsilon & \text{if } (\gamma, \theta) = (\bar{\gamma}, \bar{\theta}), \\
\nu_T(\gamma, \theta) & \text{otherwise,}
\end{cases}
\]
where we assume that \( \varepsilon \) is sufficiently small that \( \hat{\nu}_T \) satisfies sequential obedience, i.e.,
\[
\sum_{\gamma \in \Gamma_i, \theta \in \Theta} \hat{\nu}_T(\gamma, \theta) d_i (a_{-i}(\gamma), \theta) > 0
\]
for all \( i \in I \). By the dominance state assumption, we can take a \( \bar{q} < 1 \) such that
\[
\bar{q} d_i (0_{-i}, \bar{\theta}) + (1 - \bar{q}) \min_{\theta \neq \bar{\theta}} d_i (0_{-i}, \theta) > 0 \quad (A.1)
\]
for all $i \in I$. Then let $\eta > 0$ be such that
\[
\frac{\epsilon}{|I|-1} + \eta \geq \hat{q} \tag{A.2}
\]
and
\[
\sum_{\gamma \in \Gamma, \theta \in \Theta} (1 - \eta)^{|I| - n(a_{-i}(\gamma)) - 1} \nu_T(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > 0 \tag{A.3}
\]
for all $i \in I$, where $n(a_{-i}(\gamma))$ is the number of players playing action 1 in the action profile $a_{-i}(\gamma)$. Now construct the type space $(T, \pi)$ as follows. For each $i \in I$, let
\[
T_i = \begin{cases} 1, 2, \ldots & \text{if } \nu_T(\Gamma_i \times \Theta) = 1, \\ 1, 2, \ldots \cup \{\infty\} & \text{otherwise.} \end{cases}
\]
Let $\pi \in \Delta(T \times \Theta)$ be given by
\[
\pi(t, \theta) = \begin{cases} (1 - \epsilon)\eta(1 - \eta)^m \nu_T(\gamma, \theta) & \text{if } t_i < \infty \text{ for some } i \in I, \text{ and there exist } m \in \mathbb{N} \\
\frac{\epsilon}{|I|-1} & \text{if } t_1 = \cdots = t_{|I|} = \infty, \\
0 & \text{otherwise}
\end{cases}
\]
for each $t = (t_i)_{i \in I} \in T$ and $\theta \in \Theta$, where
\[
\ell(i, \gamma) = \begin{cases} \ell & \text{if there exists } \ell \in \{1, \ldots, k\} \text{ such that } i_\ell = i, \\
\infty & \text{otherwise}
\end{cases}
\]
for each $i \in I$ and $\gamma = (i_1, \ldots, i_k) \in \Gamma$. Observe that $\pi$ is consistent with $\mu$: $\sum_{\ell} \pi(t, \theta) = \sum_{\gamma} \nu_T(\gamma, \theta) = \mu(\theta)$ for all $\theta \in \Theta$.

**Claim A.1.** For any $i \in I$ and any $\tau \in \{1, \ldots, |I| - 1\}$, $\pi(\bar{\theta}|t_i = \tau) \geq \bar{q}$.

**Proof.** For $\tau \in \{1, \ldots, |I| - 1\}$, we have
\[
\pi(\bar{\theta}|t_i = \tau) = \frac{\sum_{t_{-i} \neq \tau} \pi(t_i = \tau, t_{-i}, \bar{\theta})}{\sum_{t_{-i}, \tau} \pi(t_i = \tau, t_{-i}, \theta)} \geq \frac{\epsilon}{|I|-1} + \eta \geq \bar{q},
\]
where the last inequality is by (A.2). \qed

For $S \subset I$, we denote by $1_S$ the action profile such that $a_i = 1$ if and only if $i \in S$.

**Claim A.2.** For any $i \in I$ and any $\tau \in \{|I|, |I| + 1, \ldots\}$, 
\[
\pi(\{j \neq i \mid t_j < \tau\} = S, \theta|t_i = \tau) = (1 - \eta)^{|I|-|S|-1} \nu_T(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = 1_S\} \times \{\theta\}) / C_i
\]
for all $S \subset I \setminus \{i\}$, where $C_i = \sum_{\ell=1}^{|I|} (1 - \eta)^{|I|-\ell} \nu_T(\{\gamma = (i_1, \ldots, i_k) \in \Gamma_i \mid i_\ell = i\} \times \Theta) > 0$. 

Proof. For \( \tau \in \{|I|, |I| + 1, \ldots \} \) and for \( S \subset I \setminus \{i\} \), we have
\[
\pi(\{j \neq i \mid t_j < \tau\} = S, \theta|t_i = \tau) = \pi(t_i = \tau, \{j \neq i \mid t_j < \tau\} = S, \theta)/\pi(t_i = \tau) = \frac{(1 - \varepsilon)\eta(1 - \eta)^{-|S| - 1}\nu_T(\{\gamma \in \Gamma_i \mid a_i(\gamma) = 1_S\} \times \{\theta\})/\pi(t_i = \tau)}{(1 - \eta)^{|I| - |S| - 1}\nu_T(\{\gamma \in \Gamma_i \mid a_i(\gamma) = 1_S\} \times \{\theta\})/C_i},
\]
as claimed. \( \square \)

Claim A.3. For any \( i \in I \) such that \( \tilde{\nu}_T(\Gamma_i \times \Theta) < 1 \),
\[
\pi(\{j \neq i \mid t_j < \infty\} = S, \theta|t_i = \infty) = \nu(1_S, \theta)/D_i
\]
for all \( S \subset I \setminus \{i\} \), where \( D_i = (1 - \varepsilon)(1 - \tilde{\nu}_T(\Gamma_i \times \Theta)) > 0 \).

Proof. For \( S \subset I \setminus \{i\} \), we have
\[
\pi(\{j \neq i \mid t_j < \infty\} = S, \theta|t_i = \infty) = \pi(t_i = \infty, \{j \neq i \mid t_j < \infty\} = S, \theta)/\pi(t_i = \infty) = \frac{(1 - \varepsilon)\tilde{\nu}_T(\{\gamma \in \Gamma_i \mid \bar{a}(\gamma) = 1_S\} \times \{\theta\})/D_i}{\nu_T(\{\gamma \in \Gamma_i \mid \bar{a}(\gamma) = 1_S\} \times \{\theta\})/D_i} = \nu(1_S, \theta)/D_i,
\]
as claimed, where \( (1 - \varepsilon)\tilde{\nu}_T(\gamma, \theta) = \nu_T(\gamma, \theta) \) whenever \( \bar{a}(\gamma) = 1_S \). \( \square \)

We are in a position to conclude the proof of Theorem 1. We first show that action 1 is uniquely rationalizable for all players of types \( t_i < \infty \). For types \( t_i \leq |I| - 1 \), action 1 is a strictly dominant action by Claim A.1 and condition (A.1). For \( \tau \geq |I| \), suppose that action 1 is uniquely rationalizable for all players of types \( t_i \leq \tau - 1 \). Then the expected payoff for a player \( i \) of type \( t_i = \tau \) from playing action 1 is no smaller than
\[
\sum_{S \subset I \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \tau\} = S, \theta|t_i = \tau) d_i(1_S, \theta) = \sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1 - \eta)^{|I| - n(a_i(\gamma)) - 1}\tilde{\nu}_T(\gamma, \theta)d_i(a_i(\gamma), \theta)/C_i > 0,
\]
where the equality is by Claim A.2 and the inequality by the “perturbed” sequential obedience condition (A.3). Therefore, action 1 is uniquely rationalizable for \( t_i = \tau \). Hence, by induction, action 1 is uniquely rationalizable for all types \( t_i < \infty \). Then for each \( i \in I \), let \( \sigma_i \) be the pure strategy such that \( \sigma_i(t_i)(1) = 1 \) if and only if \( t_i < \infty \). For a player \( i \) (with \( \tilde{\nu}_T(\Gamma_i \times \Theta) < 1 \)) of type \( t_i = \infty \), against \( \sigma_{-i} \) the expected payoff is given
by
\[
\sum_{S \subseteq \Gamma \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \infty\}) = S, \theta | t_i = \infty) d_i(1_S, \theta)
\]
\[
= \sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu((0, a_{-i}), \theta) d_i(a_{-i}, \theta) / D_i \leq 0,
\]
where the equality is by Claim A.3 and the inequality by (lower) obedience, which implies that playing 0 is a best response to $\sigma_{-i}$. It therefore follows that $\sigma$ is indeed the smallest equilibrium. Finally, by construction, $\sigma$ induces $\nu$, as desired.

A.1.3. Proof of Corollary 1. The “only if” part follows from Theorem 1(1) by a continuity argument. To prove the “if” part, let $\nu \in \Delta(A \times \Theta)$ satisfy consistency, obedience, and weak sequential obedience with $\nu_T \in \Delta(\Gamma \times \Theta)$. Let $\nu \in \Delta(A \times \Theta)$ be any outcome that satisfies consistency, obedience, sequential obedience with, say, $\nu_T \in \Delta(\Gamma \times \Theta)$, and grain of dominance. Then define $\nu^\varepsilon \in \Delta(A \times \Theta)$ by $\nu^\varepsilon = (1 - \varepsilon)\nu + \varepsilon \nu_T$. Clearly, $\nu^\varepsilon$ satisfies consistency, obedience, sequential obedience with $(1 - \varepsilon)\nu_T + \varepsilon \nu_T$, and grain of dominance. Hence, we have $\nu^\varepsilon \in SI$ by Theorem 1(2). Since $\nu^\varepsilon \to \nu$ as $\varepsilon \to 0$, we therefore have $\nu \in SI$.

A.1.4. Proof of Corollary 2. Part (1): For $\nu \in \Delta(A \times \Theta)$ that satisfies consistency, sequential obedience, and grain of dominance, consider the information structure as constructed in the proof of Theorem 1(2). There, all types $t_i < \infty$ of any player $i$ play action 1 as a unique rationalizable action, and hence the smallest equilibrium induces an outcome $\hat{\nu} \in SI$ that first-order stochastically dominates $\nu$.

Part (2): Let $\nu \in \Delta(A \times \Theta)$ satisfy consistency and weak sequential obedience. Then, as in the proof of Corollary 1, there exists a sequence of outcomes $\nu^\varepsilon \in \Delta(A \times \Theta)$ converging to $\nu$ that satisfy consistency, sequential obedience, and grain of dominance: for example, let $\nu^\varepsilon = (1 - \varepsilon)\nu + \varepsilon \nu_T$ with an outcome $\nu_T$ as in the proof of Corollary 1. By part (2), for each $\varepsilon$, there exists an outcome $\hat{\nu}^\varepsilon \in SI$ that first-order stochastically dominates $\nu^\varepsilon$. Then a limit point of $\hat{\nu}^\varepsilon$, which is contained in $SI$, first-order stochastically dominates $\nu$.

A.2. Proof of Proposition 2. Given any $\nu \in \Delta(A \times \Theta)$, let $N_T(\nu) = \{\nu_T \in \Delta(\Gamma \times \Theta) \mid \sum_{\gamma : \bar{a}(\gamma) = a} \nu_T(\gamma, \theta) = \nu(a, \theta)\}$ and $\Lambda(\nu) = \{\lambda \in \Delta(I) \mid \sum_{i \in I(\nu)} \lambda_i = 1\}$, which are each

\[20\]For example, let $\nu_T$ be the outcome induced by the smallest equilibrium of the information structure such that each $\theta \in \Theta$ when realized becomes common knowledge; that outcome satisfies consistency, obedience, and sequential obedience by Theorem 1(1), and grain of dominance by the dominance state assumption.
convex and compact. For $\nu_T \in N_T(\nu)$ and $\lambda \in \Lambda(\nu)$, let

$$D(\nu_T, \lambda) = \sum_{i \in I} \lambda_i \sum_{\gamma \in \Gamma, \theta \in \Theta} \nu_T(\gamma, \theta) d_i(a_{-i}(\gamma), \theta)$$

$$= \sum_{\gamma \in \Gamma, \theta \in \Theta} \nu_T(\gamma, \theta) \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta)$$

$$= \sum_{a \in A, \theta \in \Theta} \sum_{\gamma : \bar{a}(\gamma) = a} \nu_T(\gamma, \theta) \sum_{i \in S(a)} \lambda_i d_i(a_{-i}(\gamma), \theta),$$

which is linear in each of $\nu_T$ and $\lambda$.

First, $\nu$ satisfies sequential obedience (resp. weak sequential obedience) if and only if there exists $\nu_T \in N_T(\nu)$ such that $D(\nu_T, \lambda) > (\text{resp.} \geq) 0$ for all $\lambda \in \Lambda(\nu)$, which in turn is equivalent to

$$\max_{\nu_T \in N_T(\nu)} \min_{\lambda \in \Lambda(\nu)} D(\nu_T, \lambda) > (\text{resp.} \geq) 0. \quad (A.4)$$

Second, (LHS of (3.5)) = $\max_{\nu_T \in N_T(\nu)} D(\nu_T, \lambda)$ for each $\lambda \in \Lambda(\nu)$. Hence, $\nu$ satisfies condition (3.5) if and only if

$$\min_{\lambda \in \Lambda(\nu)} \max_{\nu_T \in N_T(\nu)} D(\nu_T, \lambda) > (\text{resp.} \geq) 0. \quad (A.5)$$

Now, by the minimax theorem, we have $\max_{\nu_T} \min_{\lambda} D(\nu_T, \lambda) = \min_{\lambda} \max_{\nu_T} D(\nu_T, \lambda)$, and therefore, (A.4) holds if and only if (A.5) holds.

### A.3. Proof of Proposition 3.

Suppose that the base game admits a potential $\Phi$. By Proposition 2, it suffices to show that $\nu \in \Delta(A \times \Theta)$ satisfies condition (3.5) in Proposition 2 if and only if it satisfies coalitional obedience (resp. weak coalitional obedience).

The “only if” part: Suppose that $\nu$ satisfies sequential obedience (resp. weak sequential obedience) and hence condition (3.5). Fix any $a \in A$ such that $S(a) \subseteq I(\nu)$. Define $(\lambda^a_i)_{i \in I} \in \mathbb{R}^I_+$ such that $\lambda^a_i > 0$ for some $i \in I(\nu)$ by $\lambda^a_i = 1$ if $i \in I \setminus S(a)$ and $\lambda^a_i = 0$ if $i \in S(a)$.

Consider any $(a', \theta) \in A \times \Theta$. By supermodularity, any sequence that maximizes $\sum_{i \in S(a')} \lambda^a_i d_i(a_{-i}(\gamma), \theta) = \sum_{i \in S(a') \setminus S(a)} d_i(a_{-i}(\gamma), \theta)$ over sequences $\gamma$ such that $\bar{a}(\gamma) = a'$ ranks all players in $S(a') \cap S(a)$ earlier than those in $S(a') \setminus S(a)$. Let $\gamma' = (i_1, \ldots, i_{|S(a')\cap S(a)|})$ be any such sequence, where $\{i_1, \ldots, i_{|S(a')\cap S(a)|}\} = S(a') \cap S(a)$. Thus we have

$$\max_{\gamma : \bar{a}(\gamma) = a'} \sum_{i \in S(a')} \lambda^a_i d_i(a_{-i}(\gamma), \theta) = \sum_{\ell = |S(a') \cap S(a)| + 1}^{\|S(a')\|} (\Phi((1, a_{-i}(\gamma')), \theta) - \Phi((0, a_{-i}(\gamma')), \theta))$$

$$= \Phi(a', \theta) - \Phi(a \wedge a', \theta).$$
Therefore, we have
\[ \Phi_\nu(1) - \Phi_\nu(a) = \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta)(\Phi(a', \theta) - \Phi(a \land a', \theta)) \]
\[ = \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta) \max_{\gamma; a(\gamma) = a'} \sum_{i \in S(a')} \lambda_i^a d_i(a_{-i}^{a'}(\gamma), \theta), \]
which is positive (resp. nonnegative) by condition (3.5).

The “if” part: Suppose that \( \nu \) satisfies coalitional obedience (resp. weak coalitional obedience). We want to show that \( \nu \) satisfies condition (3.5). Fix any \((\lambda_i)_{i \in I} \in \mathbb{R}_+^I\) such that \( \lambda_i > 0 \) for some \( i \in I(\nu) \). Let \( \gamma^\lambda = (i_1, \ldots, i_{|I|}) \) be a permutation of all players such that \( \{i_1, \ldots, i_{|I(\nu)|}\} = I(\nu) \) and \( \lambda_i \leq \cdots \leq \lambda_{i_{|I(\nu)|}} \). Then we have

(LHS of (3.5))
\[ \geq \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta) \sum_{i \in S(a')} \lambda_i \Phi((1, a_{-i}(\gamma^\lambda)), \theta) - \Phi((0, a_{-i}(\gamma^\lambda)), \theta)) \]
\[ = \sum_{i \in I} \lambda_i \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta) \Phi((1, a_{-i}(\gamma^\lambda)) \land a', \theta) - \Phi((0, a_{-i}(\gamma^\lambda)) \land a', \theta)) \]
\[ = \sum_{i \in I} \lambda_i (\Phi_\nu(1, a_{-i}(\gamma^\lambda)) - \Phi_\nu(0, a_{-i}(\gamma^\lambda))) \]
\[ = \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}}) \sum_{k'=k}^{|I|} (\Phi_\nu(1, a_{-i_k(\gamma^\lambda)}) - \Phi_\nu(0, a_{-i_k(\gamma^\lambda)}))) \]
\[ = \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}})(\Phi_\nu(1) - \Phi_\nu(1_{\{i_1, \ldots, i_{k-1}\}})), \]
which is positive (resp. nonnegative) by coalitional obedience (resp. weak coalitional obedience), as desired, where we set \( \lambda_{i_0} = 0 \).

A.4. Proof of Proposition 4. By Proposition 3, sequential obedience (resp. weak sequential obedience) is equivalent to coalitional obedience (resp. weak coalitional obedience) in a potential game. The “only if” part is obvious. The “if” direction follows from convexity of \( \Phi \) since for a perfect coordination outcome \( \nu \), we have
\[ \Phi_\nu(1) - \Phi_\nu(a) = \sum_{\theta \in \Theta} \nu(1, \theta)(\Phi(1, \theta) - \Phi(a, \theta)) \]
\[ \geq \left(1 - \frac{n(a)}{|I|}\right) \sum_{\theta \in \Theta} \nu(1, \theta) \Phi(1, \theta) \]
\[ = \left(1 - \frac{n(a)}{|I|}\right) \Phi_\nu(1) > (\text{resp. } \geq) 0 \]
for any \( a \neq 1 \).
A.5. **Proof of Theorem 2.** Suppose that $\Phi$ satisfies convexity and $V$ satisfies restricted convexity with respect to $\Phi$. As already noted, $\nu^*$ satisfies consistency, weak grand coalitional obedience, and obedience, and hence is in $\mathcal{ST}$.

First, we show that $(\nu^*(1, \theta))_{\theta \in \Theta}$ is an optimal solution to the problem (4.7). Let $(\nu^*(1, \theta))_{\theta \in \Theta}$ be such that $0 \leq \nu^*(1, \theta) \leq \mu(\theta)$ and $\sum_{\theta \in \Theta} \nu^*(1, \theta)V(1, \theta) < \sum_{\theta \in \Theta} \nu(1, \theta)V(1, \theta)$. Define $\xi = (\xi(\theta))_{\theta \in \Theta}$, $\xi^* = (\xi^*(\theta))_{\theta \in \Theta}$, and $\xi^{**} = (\xi^{**}(\theta))_{\theta \in \Theta}$ by $\xi(\theta) = \nu^*(1, \theta)V(1, \theta)$ for all $\theta \in \Theta$, $\xi^*(\theta) = \nu^*(1, \theta)V(1, \theta)$ for all $\theta \in \Theta$, and $\xi^{**}(\theta^* - 1) = \xi^*(\theta^* - 1) + \sum_{\theta \in \Theta} \nu^*(1, \theta)V(1, \theta) - \sum_{\theta \in \Theta} \nu^*(1, \theta)V(1, \theta) > \xi^*(\theta^* - 1)$ and $\xi^{**}(\theta) = \xi^*(\theta)$ for all $\theta \neq \theta^* - 1$.

Since $\xi^{**}$ first-order stochastically dominates $\xi$ by the construction of $\nu^*$ and $\frac{\Phi(1, \theta)}{V(1, \theta)}$ is increasing in $\theta$, we have

$$\sum_{\theta \in \Theta} \nu(1, \theta)\Phi(1, \theta) = \sum_{\theta \in \Theta} \xi(\theta)\frac{\Phi(1, \theta)}{V(1, \theta)} \leq \sum_{\theta \in \Theta} \xi^{**}(\theta)\frac{\Phi(1, \theta)}{V(1, \theta)}.$$

But we have

$$\sum_{\theta \in \Theta} \xi^{**}(\theta)\frac{\Phi(1, \theta)}{V(1, \theta)} = \sum_{\theta \in \Theta} \xi^*(\theta)\frac{\Phi(1, \theta)}{V(1, \theta)} + (\xi^{**}(\theta^* - 1) - \xi^*(\theta^* - 1))\frac{\Phi(1, \theta^* - 1)}{V(1, \theta^* - 1)} < 0,$$

since the first term equals 0 by (4.11), and $\Phi(1, \theta^* - 1) < 0$. This means that $(\nu^*(1, \theta))_{\theta \in \Theta}$ is not feasible. This implies that $(\nu^*(1, \theta))_{\theta \in \Theta}$ is an optimal solution to the problem (4.7).

Next, we show that $\nu^*$ is an optimal outcome of the adversarial information design problem. For this, it suffices to show that for any outcome $\nu \in \mathcal{ST}$, there exists a perfectly coordinated outcome $\nu'$ that satisfies consistency and weak grand coalitional obedience and whose value is no smaller than that of $\nu$. For each $(a, \theta)$, define $\alpha(a, \theta) \in [0, 1]$ by

$$\alpha(a, \theta) = \begin{cases} 1 & \text{if } \Phi(a, \theta) \leq \Phi(1, \theta), \\ \frac{n(a)}{|I|} & \text{if } \Phi(a, \theta) > \Phi(1, \theta). \end{cases}$$

Then for all $(a, \theta)$, we have $\Phi(a, \theta) \leq \alpha(a, \theta)\Phi(1, \theta)$ (by convexity) and $V(a, \theta) \leq \alpha(a, \theta)V(1, \theta)$ (by monotonicity and restricted convexity).

Take any $\nu \in \mathcal{ST}$. By Corollary 1 and Proposition 3, $\nu$ satisfies consistency and weak coalitional obedience. Define $\nu' \in \Delta(A \times \Theta)$ by

$$\nu'(a, \theta) = \begin{cases} \sum_{a' \in A}(1 - \alpha(a', \theta))\nu(a', \theta) & \text{if } a = 0, \\ \sum_{a' \in A} \alpha(a', \theta)\nu(a', \theta) & \text{if } a = 1, \\ 0 & \text{if } a \neq 0, 1, \end{cases}$$

41
which satisfies the perfect coordination property. Since \( \nu \) is consistent with \( \mu \), so is \( \nu' \).

Since \( \nu \) satisfies weak coalitional obedience, we also have

\[
\Phi_{\nu'}(1) = \sum_{\theta \in \Theta} \nu'(1, \theta) \Phi(1, \theta)
= \sum_{a \in A, \theta \in \Theta} \alpha(a, \theta) \nu(a, \theta) \Phi(1, \theta)
\geq \sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \Phi(a, \theta) = \Phi_{\nu}(1) \geq 0.
\]

Therefore, \( \nu' \) satisfies weak grand coalitional obedience. For the value of the objective function, we have

\[
\sum_{a \in A, \theta \in \Theta} \nu'(a, \theta) V(a, \theta) = \sum_{\theta \in \Theta} \nu'(1, \theta) V(1, \theta)
= \sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \alpha(a, \theta) V(1, \theta)
\geq \sum_{a \in A, \theta \in \Theta} \nu(a, \theta) V(a, \theta).
\]

This completes the proof of Theorem 2.

REFERENCES


B.1. Proof of Theorem 3(2). Let $\nu \in \Delta(A \times \Theta)$ satisfy consistency, sequential obedience, reverse sequential obedience, and two-sided grain of dominance, and let $\nu^+_I \in \Delta(\Gamma \times \Theta)$ and $\nu^-_I \in \Delta(\Gamma \times \Theta)$ be ordered outcomes establishing sequential obedience and reverse sequential obedience, respectively. By two-sided grain of dominance, there exist $\pi, \gamma$ containing all players such that $\nu^+_I(\pi, \theta) > 0$ and $\nu^-_I(\gamma, \theta) > 0$ (where $\nu^+_I(\pi, \theta) \leq \nu^+_I(0, \theta)$ and $\nu^-_I(\gamma, \theta) \leq \nu^-_I(0, \theta)$). For $\varepsilon > 0$ with $\varepsilon < \min\{\nu^+_I(\pi, \theta), \nu^-_I(\gamma, \theta)\}$, define $\tilde{\nu}^+_I, \tilde{\nu}^-_I \in \Delta(\Gamma \times \Theta)$ by

$$
\tilde{\nu}^+_I(\gamma, \theta) = \begin{cases} 
\frac{\nu^+_I(\gamma, \theta) - \varepsilon}{1 - 2\varepsilon} & \text{if } (\gamma, \theta) = (\pi, \theta), \\
\nu^+_I(\gamma, \theta) & \text{otherwise},
\end{cases}
$$

and

$$
\tilde{\nu}^-_I(\gamma, \theta) = \begin{cases} 
\frac{\nu^-_I(\gamma, \theta) - \varepsilon}{1 - 2\varepsilon} & \text{if } (\gamma, \theta) = (\gamma, \theta), \\
\nu^-_I(\gamma, \theta) & \text{otherwise},
\end{cases}
$$

where we assume that $\varepsilon$ is sufficiently small that $\tilde{\nu}^+_I$ and $\tilde{\nu}^-_I$ satisfy sequential obedience and reverse sequential obedience, respectively, i.e.,

$$
\sum_{\gamma \in \Gamma, \theta \in \Theta} \tilde{\nu}^+_I(\gamma, \theta)d_i(a_{-i}(\gamma), \theta) > 0
$$

for all $i \in I$ such that $\tilde{\nu}^+_I(\Gamma_i \times \Theta) > 0$, and

$$
\sum_{\gamma \in \Gamma, \theta \in \Theta} \tilde{\nu}^-_I(\gamma, \theta)d_i(a_{-i}^0(\gamma), \theta) < 0
$$

for all $i \in I$ such that $\tilde{\nu}^-_I(\Gamma_i \times \Theta) > 0$. Define also $\tilde{\nu} \in \Delta(A \times \Theta)$ by

$$
\tilde{\nu}(a, \theta) = \begin{cases} 
\frac{\nu(a, \theta) - \varepsilon}{1 - 2\varepsilon} & \text{if } (a, \theta) = (1, \theta), \\
\nu(a, \theta) & \text{otherwise}.
\end{cases}
$$

Observe that $\sum_{\gamma^+: a(\gamma^+)=a} \tilde{\nu}^+_I(\gamma^+, \theta) = \sum_{\gamma^-: a(\gamma^-)=a} \tilde{\nu}^-_I(\gamma^-, \theta) = \tilde{\nu}(a, \theta)$ for all $(a, \theta) \in A \times \Theta$.

By the dominance state assumption, we can take a $\tilde{q} < 1$ such that

$$
\tilde{q}d_i(0_{-i}, \theta) + (1 - \tilde{q}) \min_{\theta \neq \theta} d_i(0_{-i}, \theta) > 0,
$$

$$
\tilde{q}d_i(1_{-i}, \theta) + (1 - \tilde{q}) \max_{\theta \neq \theta} d_i(1_{-i}, \theta) < 0
$$

for all $i \in I$.
for all $i \in I$. Then let $\eta > 0$ be such that
\[
\frac{\bar{q} - 1}{\bar{q} - 1 + \eta} \geq \bar{q},
\]
and
\[
\sum_{\gamma \in \Gamma, \theta \in \Theta} (1 - \eta)|I| - n(a_{-i}(\gamma)) - 1\tilde{\nu}^+_T(\gamma, \theta)d_i(a_{-i}(\gamma), \theta) > 0
\]
for all $i \in I$ such that $\tilde{\nu}^+_T(\Gamma_i \times \Theta) > 0$, and
\[
\sum_{\gamma \in \Gamma, \theta \in \Theta} (1 - \eta)|I| - n(0, (\gamma)) - 1\tilde{\nu}^-_T(\gamma, \theta)d_i(0, (\gamma), \theta) < 0
\]
for all $i \in I$ such that $\tilde{\nu}^-_T(\Gamma_i \times \Theta) > 0$, where $n^0(a_{-i}(\gamma))$ is the number of players playing action 0 in the action profile $a_{-i}(\gamma)$.

Now construct the type space $(T, \pi)$ as follows. For each $i \in I$, let $T_i = \{1, 2, \ldots, g\}$. Define $\pi \in \Delta(T \times \Theta)$ by the following: for each $t = (s_i, a_i)_{i \in I} \in T$ and $\theta \in \Theta$, let
\[
\pi(t, \theta) = \begin{cases} 
(1 - 2\varepsilon)(1 - \eta)^m \frac{\tilde{\nu}^+_T(\gamma^+, \theta)\tilde{\nu}^-_T(\gamma^-, \theta)}{\tilde{\nu}(a, \theta)} & \text{if } \tilde{\nu}(a, \theta) > 0 \text{ and there exist } m \in \mathbb{N} \text{ and } \gamma^+, \gamma^- \in \Gamma \text{ such that } s_i = m + \ell(i, \gamma^+) \text{ for all } i \in S(a) \text{ and } s_i = m + \ell(i, \gamma^-) \text{ for all } i \in I \setminus S(a), \\
\frac{\varepsilon}{|I| - 1} & \text{if } 1 \leq s_1 = \cdots = s_{|I|} \leq |I| - 1 \text{ and } (a, \theta) = (1, \theta), (0, \theta), \\
0 & \text{otherwise}
\end{cases}
\]
where $\ell(i, \gamma) = \ell$ if $i = i_\ell$. Observe that $\pi$ is consistent with $\mu$: $\sum \pi(t, \theta) = \mu(\theta)$ for all $\theta \in \Theta$.

The rest of the proof is completed by mimicking the proof of Theorem 1(2). A similar argument as in the proof of Theorem 1(2) shows that action 1 (resp. 0) is uniquely rationalizable for all players of types $t_i = (s_i, a_i)$ with $a_i = 1$ (resp. $a_i = 0$). By construction, the unique rationalizable strategy profile, hence the unique equilibrium, induces $\nu$, as desired.

B.2. Proof of Proposition 6. We first show that any complete information BAS game has an action profile that satisfies weak sequential obedience and reverse sequential obedience. Given a complete information BAS game, let $d_i(a_{-i})$ represent the payoff increment for $a_i = 1$ over $a_i = 0$ against $a_{-i}$. For $S \subseteq I$, let $\Pi(S) \subseteq \Gamma$ denote the set of all permutations of players in $S$. An action profile $a \in A$ satisfies sequential obedience (resp.
weak sequential obedience) in the complete information BAS game \((d_i)_{i \in I}\) if there exists \(\rho \in \Delta(\Gamma)\) such that \(\rho(\Pi(S(a))) = 1\) and
\[
\sum_{\gamma \in \Pi(S(a))} \rho(\gamma) d_i(a_{-i}(\gamma)) > \text{ (resp. } \geq \text{)} 0
\]
for all \(i \in S(a)\), and it satisfies reverse sequential obedience in \((d_i)_{i \in I}\) if there exists \(\rho^0 \in \Delta(\Gamma)\) such that \(\rho^0(\Pi(I \setminus S(a))) = 1\) and
\[
\sum_{\gamma^0 \in \Pi(I \setminus S(a))} \rho^0(\gamma^0) d_i(a^0_{-i}(\gamma^0)) < 0
\]
for all \(i \in I \setminus S(a)\), where \(S(a) = \{i \in I \mid a_i = 1\}\).

**Lemma B.1.** In any complete information BAS game, there exists an action profile that satisfies weak sequential obedience and reverse sequential obedience.

To prove this, we embed the complete information BAS game \((d_i)_{i \in I}\) into a global game. Specifically, we employ a simplified global game as introduced by Frankel et al. (2003) (FMP, henceforth). Given \((d_i)_{i \in I}\), we consider the simplified global game \(G(\kappa)\) defined as follows. A state of world \(\omega\) is drawn from the real line according to the uniform distribution \(\phi\) over some large interval containing \([\omega - \kappa, \omega + \kappa]\), where \(\omega < 0\) and \(\omega > 0\) are large enough in magnitude that \(d_i(a_{-i}) + \omega < 0\) and \(d_i(a_{-i}) + \omega > 0\) for all \(i \in I\) and all \(a_{-i} \in A_{-i}\). Each player \(i \in I\) observes a noisy signal \(x_i = \omega + \kappa \varepsilon_i\), where \(\kappa > 0\) is a scale parameter, and each \(\varepsilon_i\) is distributed independently of \(\omega\) and \(\varepsilon_j, j \neq i\), according to a continuous density \(f_i\) with support contained in \([-\frac{1}{2}, \frac{1}{2}]\). For each \(i \in I\), define the function \(\tilde{d}_i : A_{-i} \times \mathbb{R} \to \mathbb{R}\) by \(\tilde{d}_i(a_{-i}, \omega) = d_i(a_{-i}) + \omega\). Player \(i\)'s payoff difference is then given by \(\tilde{d}_i(a_{-i}, x_i)\) depending directly on his signal \(x_i\). Note that these functions \((\tilde{d}_i)_{i \in I}\) satisfy the assumptions in FMP, namely, Strategic complementarities (A1), Dominance regions (A2), State monotonicity (A3), and Payoff continuity (A4). By construction, \((d_i)_{i \in I}\) is embedded in \(G^*(\kappa)\) at \(\omega = 0\).

Applied in our binary-action setting, Lemmas A1 and A4 in FMP imply the following:

- For each \(\kappa > 0\), \(G^*(\kappa)\) has an essentially unique strategy profile \((s^\kappa_i)_{i \in I}\) that survives iterative strict dominance, where \(s^\kappa_i\) is a cutoff strategy; we will identify \(s^\kappa_i\) with its cutoff.
- \((s^\kappa_i)_{i \in I}\) converges, say, to \((s_i)_{i \in I}\) as \(\kappa \to 0\).
Let $S^* = \{ i \in I \mid s_i \leq 0 \}$. Thus, $1_{S^*}$ is the action profile that the right-continuous version of the limit equilibrium strategy profile $(s_i)_{i \in I}$ plays at $x_i = 0$. We claim the following,\(^{21}\) from which Lemma B.1 follows immediately.

**Lemma B.2.** $1_{S^*}$ satisfies weak sequential obedience and reverse sequential obedience in $(d_i)_{i \in I}$.

**Proof.** Case (i): Suppose that $s_i < 0$ for all $i \in S^*$. Let $\delta > 0$ be such that $s_i \leq -\delta$ for all $i \in S^*$ and $\delta \leq s_i$ for all $i \in I \setminus S^*$. Let $\kappa \in (0, \delta]$ be such that $s_i^\kappa < -\delta/2$ for all $i \in S^*$ and $s_i^\kappa > \delta/2$ for all $i \in I \setminus S^*$. Then by construction, for any $i \in S^*$ and $j \in I \setminus S^*$, we have $s_j^\kappa - s_i^\kappa > \delta \geq \kappa$, so that player $i$ (resp. $j$) observing a signal $x_i = s_i^\kappa$ (resp. $x_j = s_j^\kappa$) knows that player $j$ (resp. $i$) receives a signal $x_j < s_j^\kappa$ (resp. $x_i > s_i^\kappa$).

Define $\rho \in \Delta(\Gamma)$ by

$$\rho(\gamma) = \mathbb{P}\left(\kappa \varepsilon_{i_1} - s_{i_1}^\kappa \geq \cdots \geq \kappa \varepsilon_{i_{|S^*|}} - s_{i_{|S^*|}}^\kappa\right)$$

for $\gamma = (i_1, \ldots, i_{|S^*|}) \in \Pi(S^*)$ and $\rho(\gamma) = 0$ for $\gamma \in \Gamma \setminus \Pi(S^*)$. Then for every player $i \in S^*$ upon observing signal $x_i = s_i^\kappa$, we have

$$0 \leq \mathbb{E}\left[\tilde{d}_i\left(1_{\{j \in S^* \setminus \{i\} \mid |x_j \geq s_j^\kappa\}}, s_i^\kappa\right) \mid x_i = s_i^\kappa\right]$$

$$= \mathbb{E}\left[\tilde{d}_i\left(1_{\{j \in S^* \setminus \{i\} \mid |x_j - s_j^\kappa \geq \kappa \varepsilon_{i_{|S^*|}} - s_{i_{|S^*|}}^\kappa\}}, s_i^\kappa\right) \mid x_i = s_i^\kappa\right]$$

$$= \mathbb{E}\left[\tilde{d}_i\left(1_{\{j \in S^* \setminus \{i\} \mid |x_j - s_j^\kappa \geq \kappa \varepsilon_{i_{|S^*|}} - s_{i_{|S^*|}}^\kappa\}}, s_i^\kappa\right)\right]$$

$$= \sum_{\gamma \in \Pi(S^*)} \rho(\gamma) d_i(a_{-i}(\gamma), s_i^\kappa),$$

where the second last equality holds due to the assumption of uniform prior. Since by state monotonicity, $\tilde{d}_i(a_{-i}, s_i^\kappa) < \tilde{d}_i(a_{-i}, 0) = d_i(a_{-i})$ for any $a_{-i}$, we have

$$\sum_{\gamma \in \Pi(S^*)} \rho(\gamma) d_i(a_{-i}(\gamma)) > 0$$

for all $i \in S^*$. Thus, $1_{S^*}$ satisfies sequential obedience with $\rho$.

By a symmetric argument, defining $\rho^0 \in \Delta(\Gamma)$ by

$$\rho^0(\gamma^0) = \mathbb{P}\left(\kappa \varepsilon_{i_1} - s_{i_1}^\kappa \leq \cdots \leq \kappa \varepsilon_{i_{|S^*|}} - s_{i_{|S^*|}}^\kappa\right)$$

for $\gamma^0 = (i_1, \ldots, i_{|S^*|}) \in \Pi(I \setminus S^*)$ and $\rho^0(\gamma^0) = 0$ for $\gamma^0 \in \Gamma \setminus \Pi(I \setminus S^*)$, we have

$$\sum_{\gamma^0 \in \Pi(I \setminus S^*)} \rho^0(\gamma^0) d_i(a_{-i}^0(\gamma^0)) < 0$$

\(^{21}\)Combined with the argument in Oyama and Takahashi (2020, Supplemental Material), the proof of Lemma B.2 shows that in BAS games, an action profile is a global game (possibly noise-dependent) selection if and only if it satisfies sequential obedience and reverse sequential obedience.
for all $i \in I \setminus S^*$. Thus, $1_{S^*}$ satisfies reverse sequential obedience with $\rho^0$.

Case (ii): Suppose that $s_i = 0$ for some $i \in S^*$. Let $\bar{\varepsilon} > 0$ be such that $s_i < \bar{\varepsilon}$ for all $i \in S^*$ and $s_i > \bar{\varepsilon}$ for all $i \in I \setminus S^*$. Then by the same argument as in case (i), for each $\varepsilon \in (0, \bar{\varepsilon}]$, there exists $\rho^\varepsilon \in \Delta(\Gamma)$ such that $\rho^\varepsilon(\Pi(S^*)) = 1$ and
\[
\sum_{\gamma \in \Pi(S^*)} \rho^\varepsilon(\gamma) \tilde{d}_i(a_{-i}(\gamma), \varepsilon) > 0
\]
for all $i \in S^*$. Let $\rho \in \Delta(\Gamma)$ be a limit point of $\rho^\varepsilon$ as $\varepsilon \to 0$, where $\rho(\Pi(S^*)) = 1$. Then by continuity, we have
\[
\sum_{\gamma \in \Pi(S^*)} \rho(\gamma) d_i(a_{-i}(\gamma)) \geq 0
\]
for all $i \in S^*$. Thus, $1_{S^*}$ satisfies weak sequential obedience with $\rho$.

Similarly, there exists $\rho^{0,\bar{\varepsilon}} \in \Delta(\Gamma)$ such that $\rho^{0,\bar{\varepsilon}}(\Pi(I \setminus S^*)) = 1$ and
\[
\sum_{\gamma^0 \in \Pi(I \setminus S^*)} \rho^{0,\bar{\varepsilon}}(\gamma^0) \tilde{d}_i(a^0_{-i}(\gamma^0), \bar{\varepsilon}) < 0
\]
for all $i \in I \setminus S^*$. By state monotonicity, we have
\[
\sum_{\gamma^0 \in \Pi(I \setminus S^*)} \rho^{0,\bar{\varepsilon}}(\gamma^0) d_i(a^0_{-i}(\gamma^0)) < 0
\]
for all $i \in I \setminus S^*$. Thus, $1_{S^*}$ satisfies reverse sequential obedience with $\rho^{0,\bar{\varepsilon}}$. \hfill $\square$

Let us go back to the original incomplete information setting. By Lemma B.2, we have the following:

**Lemma B.3.** If $\nu \in \Delta(A \times \Theta)$ satisfies consistency, sequential obedience, and two-sided grain of dominance, then there exists $\hat{\nu} \in \Delta(A \times \Theta)$ that first-order stochastically dominates $\nu$ and satisfies consistency, sequential obedience, reverse sequential obedience, and two-sided grain of dominance.

**Proof.** For each $S \subseteq I$ and $\theta \in \Theta$, apply Lemma B.2 to the complete information game $(d_i((, 1_{I \setminus S}), \theta))_{i \in S}$: let $a^*_{S,\theta} \in \prod_{i \in S} A_i$ be an action profile that satisfies weak sequential obedience and reverse sequential obedience in $(d_i((, 1_{I \setminus S}), \theta))_{i \in S}$ with $\rho_{S,\theta}$ and $\rho^0_{S,\theta}$, respectively, where $\rho_{S,\theta}(\Pi(S(a^*_{S,\theta}))) = 1$ and $\rho^0_{S,\theta}(\Pi(I \setminus S(a^*_{S,\theta}))) = 1$, and, by convention, $\rho_{\emptyset,\theta}(\emptyset) = \rho^0_{\emptyset,\theta}(\emptyset) = 1$. By construction, for any $S \subseteq I$ and $\theta \in \Theta$, we have
\[
\sum_{\gamma \in \Gamma(S) \cap \Gamma_i} \rho_{S,\theta}(\gamma) d_i(a_{-i}(\gamma'), \theta) \geq 0 \quad \text{(B.1)}
\]
for all $\gamma' \in \Pi(I \setminus S)$ and all $i \in S(a^*_{S,\theta})$, and
\[
\sum_{\gamma^0 \in \Gamma(S) \cap \Gamma_i} \rho^0_{S,\theta}(\gamma^0) d_i(a^0_{-i}(\gamma^0), \theta) < 0 \quad \text{(B.2)}
\]
for all \(i \in S \setminus S(a^*_{I,2})\), where \(\Gamma(S)\) denotes the set of sequences of distinct players in \(S\). Note, in particular, that \(a^*_{I,2} = 0\) and hence \(\rho_{I,2}(\emptyset) = 1\) by the dominance state assumption.

Let \(\nu \in \Delta(A \times \Theta)\) satisfy consistency, sequential obedience with \(\nu_{\Gamma} \in \Delta(\Gamma \times \Theta)\), and two-sided grain of dominance. Define \(\hat{\nu}_{\Gamma}, \nu^0_{\Gamma} \in \Delta(\Gamma \times \Theta)\) by

\[
\hat{\nu}_{\Gamma}(\gamma, \theta) = \sum_{\gamma' : \gamma' \prec_{\gamma'} \gamma} \nu_{\Gamma}(\gamma', \theta) \rho_{\Gamma \setminus S(\gamma'), \theta}(\gamma')
\]

and

\[
\nu^0_{\Gamma}(\gamma', \theta) = \sum_{a : S(a) \subseteq \Gamma \setminus S(\gamma')} \nu(a, \theta) \rho^0_{\Gamma \setminus S(a), \theta}(\gamma'),
\]

where for \(\gamma \in \Gamma\), \(S(\gamma)\) denotes the set of players that appear in \(\gamma\). Observe that \(\hat{\nu}_{\Gamma}(\gamma, \theta) > 0\) for some \(\gamma \in \Pi(I)\) and \(\hat{\nu}_{\Gamma}(\theta, \theta) > 0\) by two-sided grain of dominance and \(\rho_{I,2}(\emptyset) = 1\). Then define \(\hat{\nu} \in \Delta(A \times \Theta)\) by

\[
\hat{\nu}(a, \theta) = \sum_{\gamma : a(\gamma) = a} \hat{\nu}_{\Gamma}(\gamma, \theta)
= \sum_{a' : S(a') \subseteq \Gamma \setminus S(\gamma'), S(a') \setminus S(a) = S(a^*_{I,2})} \nu(a', \theta).
\]

One can verify that \(\hat{\nu}\) satisfies consistency and two-sided grain of dominance and first-order stochastically dominates \(\nu\), and that \(\nu^0_{\Gamma}\) reverse induces \(\hat{\nu}\).

Then, \(\hat{\nu}_{\Gamma}\) satisfies sequential obedience, since for each \(i \in I\), where \(\nu_{\Gamma}(\Gamma_i \times \Theta) > 0\) and \(\hat{\nu}_{\Gamma}(\Gamma_i \times \Theta) > 0\), we have

\[
\sum_{\gamma \in \Gamma, \theta \in \Theta} \hat{\nu}_{\Gamma}(\gamma, \theta) d_i(a_{-i}(\gamma), \theta)
= \sum_{\theta \in \Theta} \sum_{\gamma' \in \Gamma_i} \nu_{\Gamma}(\gamma', \theta) d_i(a_{-i}(\gamma'), \theta)
+ \sum_{\theta \in \Theta} \sum_{S \subseteq I \setminus \{i\}} \sum_{\gamma' \in \Pi(S)} \nu_{\Gamma}(\gamma', \theta) \sum_{\gamma'' \in \Gamma \setminus S(\gamma') \cap \Gamma_i} \rho_{\Gamma \setminus S, \theta}(\gamma'') d_i(a_{-i}(\gamma', \gamma''), \theta) > 0,
\]

where the inequality follows from the sequential obedience of \(\nu_{\Gamma}\) and \((B.1)\).

Finally, \(\nu^0_{\Gamma}\) satisfies reverse sequential obedience, since for each \(i \in I\), where \(\nu^0_{\Gamma}(\Gamma_i \times \Theta) > 0\), we have

\[
\sum_{\gamma \in \Gamma, \theta \in \Theta} \nu^0_{\Gamma}(\gamma, \theta) d_i(a^0_{-i}(\gamma), \theta)
= \sum_{\theta \in \Theta} \sum_{S \subseteq I \setminus \{i\}} \sum_{\gamma' \in \Pi(S)} \nu_{\Gamma}(\gamma', \theta) \sum_{\gamma^0 \in \Gamma \setminus S(\gamma') \cap \Gamma_i} \rho^0_{\Gamma \setminus S, \theta}(\gamma^0) d_i(a^0_{-i}(\gamma), \theta) > 0,
\]

where the inequality follows from \((B.2)\).
From Lemma B.3, we have the following:

**Lemma B.4.** (1) If an outcome \( \nu \) satisfies consistency, sequential obedience, and two-sided grain of dominance, then there exists an outcome \( \hat{\nu} \in FI \) that first-order stochastically dominates \( \nu \).

(2) If an outcome \( \nu \) satisfies consistency and weak sequential obedience, then there exists an outcome \( \hat{\nu} \in F \) that first-order stochastically dominates \( \nu \).

**Proof.** Part (1): Follows from Lemma B.3 and Theorem 3(2).

Part (2): Let \( \nu \in \Delta(A \times \Theta) \) satisfy consistency and weak sequential obedience. Then, as in the proof of Corollary 1, there exists a sequence of outcomes \( \nu^\varepsilon \in \Delta(A \times \Theta) \) converging to \( \nu \) that satisfy consistency, sequential obedience, and two-sided grain of dominance. By part (1), for each \( \varepsilon \), there exists an outcome \( \hat{\nu}^\varepsilon \in FI \) that first-order stochastically dominates \( \nu^\varepsilon \). Then a limit point of \( \hat{\nu}^\varepsilon \), which is contained in \( \overline{FI} \), first-order stochastically dominates \( \nu \). \( \Box \)

Finally, Proposition 6 follows from Corollary 1 and Lemma (2).