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). 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. We then characterize the optimal outcome induced by an information designer who prefers the high action to be played, but anticipates that the worst (hence smallest) equilibrium will be played. In a potential game, under convexity assumptions on the potential and the designer's objective, it is optimal to choose an outcome where actions are perfectly coordinated (all players choose the same action), with the high action played on the largest event where that action profile maximizes the average potential.

(S. Morris) Department of Economics, Massachusetts Institute of Technology

(D. Oyama) Faculty of Economics, University of Tokyo

(S. Takahashi) Faculty of Economics, University of Tokyo; Department of Economics, National University of Singapore

E-mail addresses: semorris@mit.edu, oyama@e.u-tokyo.ac.jp, satorut@e.u-tokyo.ac.jp.

Date: November 12, 2020; this version: August 10, 2022.

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

We are grateful for the comments from Gabriel Carroll, Roberto Corrao, Marina Halac, Fei Li, Elliot Lipnowski, Laurent Mathevet, Alessandro Pavan, Jacopo Perego, Chris Sandmann, Ilya Segal, Rafael Veiel, Alexander Wolitzky, Junjie Zhou, a coeditor, and four anonymous referees as well as those from conference/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, the One World Mathematical Game Theory Seminar, and SET: Seminars in Economic Theory, and at Waseda, Keio, Tsukuba, Yokohama National, Kobe, Osaka, Singapore Management, Harvard/MIT, Caltech, Kansai, Seoul National, Hitotsubashi, Cornell, Columbia, California Davis, Kansas, Michigan, Toronto, and Bar Ilan Universities. Stephen Morris gratefully acknowledges financial support from NSF Grant SES-2049744 and SES-1824137. Daisuke Oyama gratefully acknowledges financial support from JSPS KAKENHI Grants 18KK0359 and 19K01556. Part of this research was conducted while Daisuke Oyama was visiting the Department of Economics, Massachusetts Institute of Technology, whose hospitality is gratefully acknowledged.

Contents

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*. We are interested in two questions: What outcomes can be implemented by information design? And which outcome will the designer choose given an objective function?

These questions have been studied in recent years under the classical partial implementation assumption that the designer can also choose 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 is partially implementable 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 the requirement that the outcome be an (incomplete information version of) correlated equilibrium.[1]

In this paper, we study how the answers to our questions change if we are interested in a more demanding notion of implementation: *smallest equilibrium implementation*. We address these questions in the context of *binary-action supermodular (BAS) games*,[2] where a smallest equilibrium will always exist.[3] Smallest equilibrium implementation requires that the outcome be induced in the smallest equilibrium under the chosen information structure. The smallest equilibrium arises if players are initially playing the low action and switch to the high action only if it is uniquely rationalizable to do so. Our first main result addresses the implementability question by providing a characterization of smallest equilibrium implementability.

[1]Bergemann and Morris (2016) call the relevant version Bayes correlated equilibrium. Bergemann 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) labeled "communication in games", with the twist that the designer is able to deliver information to the players without having to elicit it.

[2]While we present our results in terms of abstract BAS games, such games have been used in applied settings, such as currency attacks (Morris and Shin (1998)), bank runs (Goldstein and Pauzner (2005)), political revolutions (Edmond (2013)), and contracting with externalities (Winter (2004)), among many others. Information design is a natural policy question in these applications.

[3]Throughout the paper, we will appeal to well known properties of supermodular games, without supplying explicit references. Milgrom and Roberts (1990) and Vives (1990) are classic references.

Our characterization is closely analogous to the obedience characterization of partial implementation. 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 recommended to play the high action in such a 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*. Under a dominance state assumption, that there exists a state at which the high action is a dominant action for all players, we show that an outcome is smallest equilibrium implementable if and (essentially) only if it satisfies sequential obedience, along with consistency and obedience. Thus the set of smallest equilibrium implementable outcomes, like the set of partially implementable outcomes, is characterized by a finite collection of linear constraints.

Our second main result addresses an optimal information design question: Which outcomes will be induced by an information designer who prefers the high action to be played but anticipates that the worst, hence smallest, equilibrium will be played?[4] The set of attainable outcomes, i.e., smallest equilibrium implementable outcomes, having been characterized by our first result, this result determines which outcomes are optimal given the objective function of the designer. It applies when the game has a potential (Monderer and Shapley (1996)), and the potential and the designer's objective are convex. The potential restriction requires that the sum of payoff gains from switching the actions of a subset of players not depend on the order in which they are switched; many games studied in applications (including asymmetric ones) are potential games. The convexity of the potential requires that asymmetries be not too large. Under these conditions, an optimal outcome is shown to satisfy *perfect coordination*: either all players choose the low action or all players choose the high action. This is true even with asymmetric payoffs. The designer has an instrumental motive to perfectly coordinate the players' actions, since it slackens incentive constraints by the convexity of the potential and thus enables the designer to induce higher outcomes. Convexity of the designer's objective, i.e., her intrinsic preference for coordination, only increases the advantages of perfect coordination. Solving our information design problem then reduces to solving a simple

[4]Assuming that the designer can select the equilibrium bypasses the issue of eliminating the possibility of coordination failure, a key issue in games with strategic complementarities.

Bayesian persuasion problem. Say that a state is “good” if the potential of all playing the high action is higher than the potential of all playing the low action (normalized to zero), and “bad” otherwise. It is then optimal to pool all the good states with as many bad states with the lowest cost benefit ratio as possible, subject to the average potential being nonnegative, where the cost of including a state is given by the loss in the potential at that state, while the benefit is the gain in the objective at that state.

1.1. **Related Literature.** Our implementation result has its roots in a large literature on the role of higher order beliefs in games. While not expressed in this language, the “electronic mail game” of Rubinstein (1989) and the global games of Carlsson and van Damme (1993) showed that the risk dominant equilibrium of a two-player two-action coordination game can be implemented by information design if there is a small probability of dominant action types. Oyama and Takahashi (2020) generalize these arguments to general BAS games, appealing to a complete information version of sequential obedience (as an intermediate step of a proof). Our argument establishing the sufficiency of sequential obedience for smallest equilibrium implementation generalizes this logic to an incomplete information setting. Kajii and Morris (1997) showed a converse: in any incomplete information setting, if payoffs are given by a fixed complete information game with high probability, there is always an equilibrium where the risk dominant equilibrium of the fixed game is played with high probability. Our argument establishing the necessity of sequential obedience for smallest equilibrium implementation relates to Kajii and Morris (1997) and later work on “robustness to incomplete information”. These results establish necessary conditions on arbitrary common prior type spaces by appealing to revelation principle like arguments.

If the outcome to be smallest equilibrium implemented has all players choose the high action, our characterization can be read as characterizing the set of games for which playing the high action is uniquely rationalizable. This interpretation of our results provides a new perspective on bilateral contracting with externalities (Segal (2003) and Winter (2004)), where agents make decisions about whether to participate or not in the presence of strategic complementarities. A recent paper of Halac et al. (2021) shows how the “divide-and-conquer” incentive schemes (Segal (2003)) in the model of Winter (2004) can be improved upon by introducing higher order payoff uncertainty. Moriya and Yamashita (2020) incorporate incomplete information into this problem. Our sequential obedience condition can be considered as a stochastic version of the "divide-and-conquer" condition and characterizes the smallest equilibrium (hence full) implementation constraint, subject to which the total transfer is minimized in these problems.

Thus, an important contribution of this paper is to provide a framework within which to identify the tight connection between smallest equilibrium implementation and the literatures on higher order beliefs in games and contracting with externalities. In Section 3.3, we formally discuss the connection by specializing to the case where a single state is assigned probability 1 or close to 1.

There is a recent small literature on information design with adversarial equilibrium selection, in particular for BAS games. Three papers are most relevant.[5] Inostroza and Pavan (2022) posed the question in the context of a class of regime change games (unlike us, they assumed that players had private information prior to the designer's information release). Mathevet et al. (2020) also posed the question and solved for an optimal information structure in a two-player two-state symmetric example. A recent paper of Li et al. (2022) solved for an optimal information structure in regime change games. Inostroza and Pavan (2022) showed that it was without loss to assume that optimal outcomes satisfied the perfect coordination property in regime change games, which also held for the optimal outcomes in Mathevet et al. (2020) and Li et al. (2022). However, all these papers assume symmetric payoffs, where the perfect coordination property is hardly surprising.[6] The perfect coordination property is much more surprising in games with asymmetric payoffs. We illustrate this point in Section 2 with an asymmetric version of the example of Mathevet et al. (2020). There is generally multiplicity in implementing information structures. The information structures implementing the optimal outcome

[5] Earlier, Kamien et al. (1990) raised the question of full implementation by information design and demonstrated by examples how private signals could generate more outcomes than public signals. Carroll (2016) considered a bilateral trading game and characterized the information structure which minimized the sum of players' payoffs, subject to the best equilibrium being played. The trading game had binary actions but was not supermodular, and the methods were different from this paper. Bergemann and Morris (2019, Section 7.1) and Hoshino (2022) illustrate the implications of the higher order beliefs literature for information design.

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 study partial implementation by the design of the information structure *and* the extensive form.

[6] Inostroza and Pavan (2022, Additional Material) show that assuming the perfect coordination property is without loss in regime change games with asymmetric payoffs for the objective of minimizing the probability of regime change (as well as under some other alternative generalized settings). Their argument is, however, special to the regime change payoffs, and neither implies nor is implied by our result. In Supplemental Appendix B.5, we show that the perfect coordination property holds in asymmetric regime change games within our setting as well.

in Mathevet et al. (2020) and Li et al. (2022) are tailored to the applications, whereas our implementing information structure construction applies to general BAS games.

The recent literature on Bayesian persuasion (Kamenica and Gentzkow (2011)) has highlighted the distinction between a belief-based modeling of incomplete information (i.e., identifying information with a probability distribution over posteriors satisfying "Bayes-plausibility") and a signal-based approach (identifying information with a mapping from states to a probability distribution over signals). The many-player analogue of the belief-based approach is to look at (common prior subspaces of) the universal type space of Mertens and Zamir (1985) (Mathevet et al. (2020) and Sandmann (2020)) that encodes players' beliefs and higher order beliefs. Our results embed restrictions on higher order beliefs imposed by the common prior assumption, and one could make this explicit, as Kajii and Morris (1997) and Oyama and Takahashi (2020) did (using the language of belief operators (Monderer and Samet (1989)) and generalized belief operators (Morris and Shin (2007) and Morris et al. (2016)), respectively). We choose not to work directly with the universal type space or explicitly with beliefs and higher order beliefs, because our sequential obedience approach is simpler, highlights the analogy with the partial implementation case, and reduces the original infinite-dimensional problem into a finite-dimensional linear program.

1.2. **Setting.** There are finitely many players, denoted by $I = \{1, \dots, |I|\}$, $|I| \geq 2$. A state is drawn from a finite set Θ according to the probability distribution $\mu \in \Delta(\Theta)$,[7] where we assume that μ 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 $i \in I$ and $\theta \in \Theta$,

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

is weakly increasing in $a_{-i} \in A_{-i}$. We denote $\mathbf{0} = (0, \dots, 0)$ and $\mathbf{1} = (1, \dots, 1)$, 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 requires that there exist a state where action 1 is a dominant action for all players: i.e., there exists $\overline{\theta} \in \Theta$

[7] For a finite or countably infinite set X, we write $\Delta(X)$ for the set of all probability distributions over X.

such that $d_i(\mathbf{0}_{-i}, \overline{\theta}) > 0$ for all $i \in I$. This is a richness assumption about the space of possible payoff structures and technically will be used to trigger an infection argument in smallest equilibrium implementation.[8] Fixing I, A, Θ, and μ, 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 $\mathcal{T} = ((T_i)_{i\in I}, \pi)$, in which each T_i is a countable set of types for player $i \in I$,[9] 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 μ: $\sum_{t\in T} \pi(t, \theta) = \mu(\theta)$ 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 $\mathcal{T}$ defines an incomplete information game, which we refer to simply as $\mathcal{T}$. In the incomplete information game $\mathcal{T}$, a strategy for player i is a mapping $\sigma_i \colon T_i \to \Delta(A_i)$. A strategy profile $\sigma = (\sigma_i)_{i\in I}$ is a (Bayes-Nash) equilibrium of the game $\mathcal{T}$ 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 $E(\mathcal{T})$ for the set of equilibria of the game $\mathcal{T}$. Since the game is supermodular, there always exists a smallest equilibrium, which is in pure strategies, and this equilibrium is also the limit of best response dynamics, if all players initially always choose action 0. We write $\underline{\sigma}(\mathcal{T})$ 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 $(\mathcal{T}, \sigma)$ 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 ν 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 what is assumed about the equilibrium to be played. Two extreme cases are studied in the mechanism design literature:

[8] This assumption will be maintained throughout the analysis and used in Theorem 1(2) (and results that use Theorem 1(2)). This form of the assumption, however, is stronger than needed. See Supplemental Appendix B.4.6 for a relaxation.

[9] The countability restriction is made for expositional simplicity only. In particular, Theorem 1(1) holds with possibly uncountable measurable spaces of types; see Supplemental Appendix B.4.5.

partial implementation requires only that some equilibrium induce the outcome, and *full implementation* requires that all equilibria induce the outcome. We will focus on an intermediate case (well defined for supermodular games): *smallest equilibrium implementation* requires that the smallest equilibrium induce the outcome.

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

An outcome ν satisfies *obedience* if

$$\sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu((a_i, a_{-i}), \theta) \left(u_i((a_i, a_{-i}), \theta) - u_i((a_i', a_{-i}), \theta)\right) \geq 0 \tag{1.1}$$

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) 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.

A more demanding notion of implementation is:

Definition 2. Outcome ν is *fully implementable* if there exists an information structure $\mathcal{T}$ such that $(\mathcal{T}, \sigma)$ induces ν for all $\sigma \in E(\mathcal{T})$.[10]

And the intermediate notion we study is:

Definition 3. Outcome ν is *smallest equilibrium implementable* (*S-implementable*) if there exists an information structure $\mathcal{T}$ such that $(\mathcal{T}, \underline{\sigma}(\mathcal{T}))$ induces ν.

We write $SI \subset \Delta(A \times \Theta)$ (resp. $FI \subset \Delta(A \times \Theta)$) for the set of S-implementable (resp. fully implementable) outcomes. Clearly, $FI \subset SI \subset BCE$. We characterize SI and its closure $\overline{SI}$ in Section 3. A characterization of FI is reported in Supplemental Appendix B.3.

Smallest equilibrium implementation is relevant for an information designer who expects the smallest equilibrium to be played. For example, Segal (2003, Section 4.1.3) discusses contracting applications where the smallest equilibrium is the Pareto-efficient

[10]Under supermodularity, full implementation in fact requires $E(\mathcal{T})$ be a singleton.

equilibrium for the players. Cooper (1994) argues for hysteresis equilibrium selection, where past actions are default choices, and players switch from the default only if it is uniquely rationalizable to do so. If action 0 was the default action, this would lead to smallest equilibrium selection. In this paper, we study the problem of an information designer who favors the high action but anticipates adversarial equilibrium selection as a worst-case scenario. We introduce this problem in the next subsection and show how this has an S-implementation characterization.

1.4. **Optimality.** Now we postulate an information designer who optimally chooses an information structure $\mathcal{T}$ based on her welfare criterion over $A \times \Theta$. Suppose that the designer receives a value $V(a, \theta)$ 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 are interested in the information design problem with adversarial equilibrium selection, where the designer wants to obtain the best possible values even if players will play her worst equilibrium, which, by the monotonicity of V in a, is the smallest equilibrium $\underline{\sigma}(\mathcal{T})$. Thus her problem is:

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

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

$$\sup_{\nu \in SI} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta) V(a, \theta) = \max_{\nu \in \overline{SI}} \sum_{a \in A, \theta \in \Theta} \nu(a, \theta) V(a, \theta). \tag{1.2}$$

An *optimal outcome* of the adversarial information design problem is any element ν of $\overline{SI}$ that maximizes $\sum_{a, \theta} \nu(a, \theta) V(a, \theta)$.

2. A Leading Example

We will use the following example to illustrate ideas throughout the paper.[11] Let us label the action 1 "invest" and the action 0 "not invest". The payoff to not invest is always 0. There are two players, $I = \{1, 2\}$. Player 1 has a cost 7 of investing while player 2 has a cost 8. Each player receives a return of 3 to investing when the other player invests, so the game is supermodular. There are two states, $\mathbf{b}$ ("bad") and $\mathbf{g}$ ("good"), which are equally likely ($\mu(\mathbf{b}) = \mu(\mathbf{g}) = \frac{1}{2}$). If the state is good, players receive an additional return 9 to investing. Thus both players have a dominant action to invest

[11] A detailed analysis of this example is given in Supplemental Appendix B.1.

in the good state and not invest in the bad state (hence, the dominance state assumption is satisfied with $\overline{\theta} = \mathbf{g}$). The payoffs are summarized by the following tables, where player 1 is the row player and player 2 is the column player:

b	Not	Invest
Not	$0, 0$	$0, -8$
Invest	$-7, 0$	$-4, -5$

g	Not	Invest
Not	$0, 0$	$0, 1$
Invest	$2, 0$	$5, 4$

. (2.1)

Consider the problem of a designer who wants to maximize the expected number of players who choose action 1 (i.e., $V(a, \theta) = |\{i \in I \mid a_i = 1\}|$ for all $a \in A$ and $\theta \in \Theta$).

First, consider the case of partial implementation. By the asymmetry of the payoffs, the optimal outcome is asymmetric (Arieli and Babichenko (2019)). The optimal direct information structure and equilibrium are the following. Player 1 (more willing to invest) is always recommended to invest (hence receives no information). Player 2 is recommended to invest always in the good state and with probability $\frac{4}{5}$ in the bad state (otherwise, recommended not to invest). To verify that following the recommendations constitutes an equilibrium, observe that player 2 is (just) willing to invest when recommended to do so since he is sure that player 1 will invest and assigns to the good state probability $\frac{5}{9}$ which is the smallest probability with which he is willing to invest. Given this, player 1 is (strictly) willing to invest. The resulting outcome (probability distribution over actions and states) is represented in the following:

b	Not	Invest
Not	0	0
Invest	$\frac{1}{10}$	$\frac{2}{5}$

g	Not	Invest
Not	0	0
Invest	0	$\frac{1}{2}$

,

where the expected number of players who invest is $\frac{19}{10}$. The optimal outcome is not a perfect coordination outcome (an outcome where either both invest or both do not invest). This is because any partially implementable perfect coordination outcome entails a slack in the obedience constraint for player 1 (more willing to invest) that can be exploited to induce more investment.

However, in the direct information structure as described, there is a strict equilibrium where both players never invest (which is the smallest equilibrium thereof): if player 1 thinks that player 2 will never invest, his expected payoff to investing is negative (and even smaller for player 2). No outcome close to the partially implementable outcome above is S-implementable.

Our Theorem 1 in Section 3 will establish that the following perfectly coordinated outcome is S-implementable (and indeed fully implementable) for all $0 < \delta \leq \frac{1}{4}$:

b	Not	Invest
Not	$\frac{1}{4} + \delta$	0
Invest	0	$\frac{1}{4} - \delta$

g	Not	Invest
Not	0	0
Invest	0	$\frac{1}{2}$

, (2.2)

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

b	Not	Invest
Not	$\frac{1}{4}$	0
Invest	0	$\frac{1}{4}$

g	Not	Invest
Not	0	0
Invest	0	$\frac{1}{2}$

. (2.3)

The expected number of players who invest is $\frac{3}{2}$ under the latter outcome. Our Theorem 2 in Section 4 will establish that this outcome is the solution to the information design problem under S-implementability.

To provide intuition for these results, suppose that players observed a public "good" signal always in the good state and with probability $\frac{1}{2}$ in the bad state (otherwise, they observe a "bad" signal). If players both observed the good signal, they would think that the state was good with probability $\frac{2}{3}$, and the expected payoffs would be

	Not	Invest
Not	$0, 0$	$0, -2$
Invest	$-1, 0$	$2, 1$

.

This "average" game has two strict Nash equilibria, (Not Invest, Not Invest) and (Invest, Invest). The (Invest, Invest) equilibrium is just risk dominant (Harsanyi and Selten (1988)); also this is a potential game (Monderer and Shapley (1996)) and (Invest, Invest) weakly maximizes the potential. To (approximately) implement (Invest, Invest) in a smallest equilibrium—hence as a unique rationalizable play—by eliminating (Not Invest, Not Invest), the direct information structure as described so far does not suffice, and we need private signals. From the higher order beliefs literature, we know that an "email game information structure" (Rubinstein (1989)) will uniquely implement a risk dominant equilibrium, if we allow for a (vanishingly small) possibility of dominant action types. In Section 3.2, we will describe such an information structure that implements the outcome (2.2), as an illustration of the proof for the general case of Theorem 1. Observe also that the probability $\frac{1}{2}$ with which the "good" signal is sent in the bad state is the largest probability such that (Invest, Invest) is risk dominant in the average game.

With a larger probability, (Not Invest, Not Invest) would become a strictly risk dominant equilibrium, which cannot be eliminated by dominant action types of small probability (Kajii and Morris (1997)). In Section 3.3, we discuss formal connections between our characterizations and the literature on higher order beliefs.[12]

This example illustrates that an optimal outcome exhibits the perfect coordination property *even in asymmetric games* for S-implementation but not for partial implementation. Note that if we had considered a symmetric game, clearly the perfect coordination property would have held for partial implementation as well. Thus, the perfect coordination results in Mathevet et al. (2020) (in a symmetric version of this example) and Li et al. (2022) (in regime change games) are hardly surprising. The perfect coordination property holds in this example despite the costs being asymmetric. We will see in Section 3.6 that this example has a convex potential (so the asymmetry is not too large). Given the perfect coordination property, it is then optimal to have the players invest on the largest probability event where, in the induced average game, (Invest, Invest) is risk dominant, or equivalently, maximizes the average potential. The arguments in Section 4 extend these ideas to the general case under convexity assumptions on the potential and the designer objective.

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 action 0 but the information designer recommended a subset of players to play action 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 Γ be the set of all sequences of distinct players. For example, if $I = \{1, 2, 3\}$, then

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

[12]Mathevet et al. (2020) analyzed a symmetric version of this example, but did not note that they were implementing both invest on the largest event where both invest was risk dominant.

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

$$\nu(a, \theta) = \sum_{\gamma : a(\gamma) = a} \nu_\Gamma(\gamma, \theta).$$

For each $i \in I$, let Γ_i be the set of all sequences in Γ 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 γ before i (therefore, player j plays action 0 if and only if either j is not listed in γ or j is listed in γ after i).[13]

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

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > 0 \tag{3.1}$$

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.

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 ν and satisfies sequential obedience (resp. weak sequential obedience).

By definition, the ordered outcome ν_Γ such that $\nu_\Gamma(\emptyset, \theta) = \mu(\theta)$ for all $\theta \in \Theta$ and hence the outcome ν such that $\nu(\mathbf{0}, \theta) = \mu(\theta)$ 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 ν_Γ given by

	b	**g**
$\emptyset$	$\frac{1}{4}$	0
1	0	0
2	0	0
12	$\frac{1}{6}$	$\frac{1}{3}$
21	$\frac{1}{12}$	$\frac{1}{6}$

.

[13]The notation $a_{-i}(\gamma)$ should not be confused with the possible notation "$(a(\gamma))_{-i}$" (which would represent the action profile of player i's opponents such that all the players listed in γ play action 1).

This satisfies weak sequential obedience:

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

$$\sum_{\gamma\in\Gamma_2,\theta\in\Theta} \nu_\Gamma(\gamma,\theta)d_2(a_{-2}(\gamma),\theta) = \frac{1}{6}\times(-5)+\frac{1}{12}\times(-8)+\frac{1}{3}\times 4+\frac{1}{6}\times 1 = 0,$$

and hence the induced outcome (2.3) also does. Reducing $\nu_\Gamma(12,\mathbf{b})$ or $\nu_\Gamma(21,\mathbf{b})$ by $\delta > 0$ and increasing $\nu_\Gamma(\emptyset,\mathbf{b})$ by δ makes these values strictly positive, and hence the resulting ordered outcome satisfies sequential obedience, as does the induced outcome (2.2). Note that, while the outcome (2.3) is symmetric (and satisfies perfect coordination), the inducing ordered outcome above treats the players asymmetrically: the asymmetry in the payoffs is absorbed in the asymmetry in the "hidden variables" $\nu_\Gamma(\gamma,\theta)$.

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 $\overline{\theta}$.

Definition 6. Outcome ν satisfies *grain of dominance* if $\nu(\mathbf{1},\overline{\theta}) > 0$.

We now state our first 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 is given in Appendix A.1, and those of its corollaries below in Supplemental Appendix B.2.

Because of the strict inequalities in the definition of sequential obedience, the set SI of S-implementable outcomes is not closed in general.[14] Its closure $\overline{SI}$, on the other hand, is cleanly characterized by weak sequential obedience (without referring to grain of dominance). In particular, analogous to BCE, $\overline{SI}$ is a convex polytope.[15]

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

[14] Also, the grain of dominance condition is indispensable for part (2); see Supplemental Appendix B.4.7 for a counter-example.

[15] 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 $\nu_\Gamma \in \Delta(\Gamma\times A)$ to $\nu \in \Delta(A\times\Theta)$ by $\nu(a,\theta) = \sum_{\gamma:a(\gamma)=a} \nu_\Gamma(\gamma,\theta)$.

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 (i.e., the condition (1.1) with $a_i = 1$ and $a_i' = 0$). If an outcome satisfies sequential obedience, but not the "lower obedience" requirement that a player want to follow a recommendation to play action 0 (i.e., the condition (1.1) with $a_i = 0$ and $a_i' = 1$), then, by the construction in the proof of Theorem 1(2), we can find a first-order stochastically dominating outcome in SI.[16] By a continuity argument, we thus have:

Corollary 2. *If an outcome ν satisfies consistency and weak sequential obedience, then there exists an outcome $\hat{\nu} \in \overline{SI}$ that first-order stochastically dominates ν.*

Corollaries 1 and 2 have important implications to the adversarial information design problem (1.2). 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, therefore, 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).

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, say by a round-robin protocol, 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. By supermodularity, the sequence of strategy profiles will

[16]For $\nu, \hat{\nu} \in \Delta(A \times \Theta)$, we say that $\hat{\nu}$ first-order stochastically dominates ν if for each $\theta \in \Theta$, $\hat{\nu}(\cdot,\theta)$ first-order stochastically dominates $\nu(\cdot,\theta)$: $\sum_{a\in B} \hat{\nu}(a,\theta) \geq \sum_{a\in B} \nu(a,\theta)$ for all upper sets $B \subset A$ (i.e., sets B such that $a' \in B$ whenever $a \in B$ and $a' \geq a$).

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 γ corresponding to the order in which those players switch. Let us define an ordered outcome by letting the probability of state θ and sequence γ be the probability that θ is the state and γ 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 realized sequence (unknown to him) to play action 1.

Second, consider sufficiency (i.e., part (2)). The proof is by construction. As we have seen in Section 2, the direct information structure does not in general S-implement the target outcome. But as known from the email game and global games that, in the presence of dominant action types, higher order uncertainty about payoffs can induce, for example, a risk dominant action or a potential maximizer to spread through infection effects. Here we describe the construction by showing how to S-implement outcome (2.2) in the example in Section 2. When $\theta = \mathbf{b}$, it is publicly announced with (ex ante) probability $\frac{1}{4} + \delta$. On the remaining event, where invest is risk dominant, private signals are sent, as in the email game or global games, in such a way that all types of both players will find invest iteratively dominant. The ordered outcome establishing sequential obedience gives a general recipe to construct such an information structure. As we noted above, the ordered outcome

	$\mathbf{b}$	$\mathbf{g}$
$\emptyset$	$\frac{1}{4}+\delta$	0
1	0	0
2	0	0
12	$\frac{1}{6}-\delta$	$\frac{1}{3}$
21	$\frac{1}{12}$	$\frac{1}{6}$

(3.2)

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

$$\left(\frac{1}{6}-\delta\right)\times(-7)+\frac{1}{12}\times(-4)+\left(\frac{1}{3}-\varepsilon\right)\times 2+\frac{1}{6}\times 5>0, \tag{3.3}$$

$$\left(\frac{1}{6}-\delta\right)\times(-5)+\frac{1}{12}\times(-8)+\left(\frac{1}{3}-\varepsilon\right)\times 4+\frac{1}{6}\times 1>0. \tag{3.4}$$

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

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

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 θ, a sequence γ of players is drawn, independently of m, according to $\nu_\Gamma(\cdot,\theta)$ in (3.2), but with $\nu_\Gamma(12,\mathbf{g})-\varepsilon$ in place of $\nu_\Gamma(12,\mathbf{g})$. If $\gamma = 12$ or 21, then each player receives a signal equal to the sum of m and his ranking in γ; if $\gamma = \emptyset$, both receive a signal ∞. The remaining probability ε is relocated to $\pi((1,1),\mathbf{g})$, 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

b

$t_1 \backslash t_2$	1	2	3	4	$\cdots$	∞
1		$\eta\left(\frac{1}{6}-\delta\right)$				
2	$\eta\frac{1}{12}$		$\eta(1-\eta)\left(\frac{1}{6}-\delta\right)$			
3		$\eta(1-\eta)\frac{1}{12}$		$\eta(1-\eta)^2\left(\frac{1}{6}-\delta\right)$		
4			$\eta(1-\eta)^2\frac{1}{12}$		$\ddots$	
$\vdots$				$\ddots$		
∞						$\frac{1}{4}+\delta$

g

$t_1 \backslash t_2$	1	2	3	4	$\cdots$	∞
1	ε	$\eta\left(\frac{1}{3}-\varepsilon\right)$				
2	$\eta\frac{1}{6}$		$\eta(1-\eta)\left(\frac{1}{3}-\varepsilon\right)$			
3		$\eta(1-\eta)\frac{1}{6}$		$\eta(1-\eta)^2\left(\frac{1}{3}-\varepsilon\right)$		
4			$\eta(1-\eta)^2\frac{1}{6}$		$\ddots$	
$\vdots$				$\ddots$		
∞						

Table 1. Information structure implementing outcome (2.2)

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. **The (Limit) Complete Information Case.** In this section, we discuss the sequential obedience condition and our characterization result for S-implementability in the special case where for some state $\theta^* \in \Theta$, we have either $\mu(\theta^*) = 1$, or $\mu(\theta^*)$ converging to 1. This allows us to establish the tight connection between our results and the literatures on contracting with externalities (Segal (2003), Winter (2004), Halac et al. (2021)) and on higher order beliefs, in particular on robustness to incomplete information (Kajii and Morris (1997), Oyama and Takahashi (2020)).

3.3.1. *(Limit) S-Implementation and Sequential Obedience.* First, suppose that we relax our maintained full support assumption for the probability distribution μ on states, and assume instead that $\mu(\theta^*) = 1$. Thus, the base game can be considered as a complete information game, and a consistent outcome, which assigns probability 1 to θ^*, can be identified with a probability distribution over action profiles $\xi \in \Delta(A)$. Let a complete information BAS game be given and represented by a profile $(f_i)_{i\in I}$ of payoff difference functions $f_i\colon A_{-i} \to \mathbb{R}$, $i \in I$. Then the set of partially implementable outcomes in $(f_i)_{i\in I}$ is equal to the set of correlated equilibria of $(f_i)_{i\in I}$. By supermodularity, there is a smallest correlated equilibrium, which is the degenerate outcome on the smallest Nash equilibrium. This is the unique S-implementable outcome, and the smallest Nash equilibrium $\underline{a}$ is reached by iterative dominance from $\mathbf{0}$ (all playing 0), i.e., there exists $\gamma \in \Gamma$ such that $a(\gamma) = \underline{a}$ and

$$f_i(a_{-i}(\gamma)) > 0 \tag{3.5}$$

for all $i \in I$ such that $\underline{a}_i = 1$. In particular, $\mathbf{1}$ (all playing 1) is S-implementable (hence fully implementable) in $(f_i)_{i\in I}$ if and only if there exists a permutation γ of all players that satisfies (3.5) for all $i \in I$.

This observation lies behind the literature on bilateral contracting with externalities (Segal (2003), Winter (2004)), where the authors consider an exogenous initial supermodular game and add transfers in some form (thus determining the payoff functions f_i endogenously) to implement a target outcome as a unique equilibrium. They then ask what is the least-cost way of providing transfers so that condition (3.5) is satisfied. We refer to (3.5) as the "divide-and-conquer" condition following the terminology in this literature.

Our sequential obedience condition can be understood as a stochastic divide-and-conquer condition, where the ordering of the players is random and the condition is written as the expectation with respect to the random ordering. Formally, an ordered outcome $\rho \in \Delta(\Gamma)$ satisfies sequential obedience (resp. weak sequential obedience) in a complete information game $(f_i)_{i\in I}$ if

$$\sum_{\gamma\in\Gamma_i} \rho(\gamma) f_i(a_{-i}(\gamma)) > \text{(resp. } \geq \text{) } 0 \tag{3.6}$$

for all $i \in I$ such that $\rho(\Gamma_i) > 0$. An outcome $\xi \in \Delta(A)$ satisfies sequential obedience (resp. weak sequential obedience) in $(f_i)_{i\in I}$ if there exists an ordered outcome $\rho \in \Delta(\Gamma)$ that induces ξ (i.e., $\xi(a) = \sum_{\gamma:a(\gamma)=a} \rho(\gamma)$ for all $a \in A$) and satisfies sequential obedience (resp. weak sequential obedience) in $(f_i)_{i\in I}$.

By Theorem 1, (the weak version of) condition (3.6) characterizes S-implementation in the limit complete information case as $\mu(\theta^*) \to 1$. For a prior $\mu \in \Delta(\Theta)$, let $SI(\mu) \subset \Delta(A \times \Theta)$ denote the set of S-implementable outcomes under μ. We say that an outcome $\xi \in \Delta(A)$ is *limit S-implementable* at θ^* if there exist a sequence of priors $\mu^k \in \Delta(\Theta)$ and a sequence of S-implementable outcomes $\nu^k \in SI(\mu^k)$ such that $\mu^k(\theta^*) \to 1$ and $\sum_{\theta\in\Theta} \nu^k(\cdot, \theta) \to \xi$ as $k \to \infty$. Under the maintained assumption of dominance state, we have the following by Theorem 1 along with an argument similar to the proofs of Corollaries 1 and 2.

Corollary 4. (1) *An outcome is limit S-implementable at θ^* if and only if it satisfies obedience and weak sequential obedience in $(d_i(\cdot, \theta^*))_{i\in I}$.*

(2) *If an outcome ν satisfies weak sequential obedience in $(d_i(\cdot, \theta^*))_{i\in I}$, then there exists an outcome $\hat{\nu}$ that is limit S-implementable at θ^* and first-order stochastically dominates ν.*

The proof is given in Supplemental Appendix B.2.3.

As an illustration, consider the case of two players, and suppose that in the complete information game at θ^*, both $\mathbf{1}$ and $\mathbf{0}$ are strict equilibria, so that $\mathbf{0}$ is the S-implementable, but not fully implementable, outcome when $\mu(\theta^*) = 1$. It can readily be verified that $\mathbf{1}$ satisfies weak sequential obedience if and only if it is weakly risk dominant. Therefore, the "if" part of Corollary 4(1) implies that a risk dominant equilibrium of a two-player two-action coordination game is limit S-implementable (hence limit fully implementable), a well-known result from the infection arguments of the email game (Rubinstein (1989)) and global game (Carlsson and van Damme (1993)).

Corollary 4(2) has been obtained by Oyama and Takahashi (2020) as an intermediate proof step in establishing a necessary condition for "robustness to incomplete information" (to be discussed below). They introduced the signal structure in which each player i receives a signal given by the sum of a nonnegative integer drawn from an almost uniform distribution and i's ranking in the sequence γ drawn from an ordered outcome $\rho \in \Delta(\Gamma)$ satisfying sequential obedience. Under this signal structure, players are shown to play action 1 as a unique rationalizable action whenever they appear in the sequence, by an infection argument starting from types for which action 1 is a dominant action. Hence, *some* outcome that stochastically dominates the outcome induced by ρ is limit S-implementable.[17] Thus, our construction in the proof of Theorem 1(2) is an incomplete information generalization of that of Oyama and Takahashi (2020) (which in turn is a generalization of the infection argument of the email game), where we need to exactly S-implement a *given* outcome that satisfies sequential obedience under a fixed prior $\mu \in \Delta(\Theta)$ (along with extra conditions).

3.3.2. *Robustness to Incomplete Information.* Corollary 4 (hence Theorem 1) is closely related to the concept of *robustness to incomplete information* (Kajii and Morris (1997)). An action profile a^* of a complete information game is robust if any nearby incomplete information game where payoffs are given by that complete information game with high probability has an equilibrium where a^* is played with high probability. In particular, a robust equilibrium is immune to infection of other equilibria. Adapted to the current framework, this concept is formally defined as follows (stated only for the particular action profile $\mathbf{0}$ in order to allow a tight connection with S-implementation). A complete information BAS game is given by a profile of payoff difference functions $f_i \colon A_{-i} \to \mathbb{R}$, $i \in I$. Fix a (finite) state space Θ and a (supermodular) base game $(d_i)_{i\in I}$ such that $d_i(\cdot, \theta^*) = f_i(\cdot)$ for all $i \in I$ while maintaining the dominance state assumption. Action profile $\mathbf{0}$ is *robust* in $(f_i)_{i\in I}$ if for any sequence of priors $\mu^k \in \Delta(\Theta)$ such that $\mu^k(\theta^*) \to 1$ as $k \to \infty$ and for any sequence of S-implementable outcomes $\nu^k \in \Delta(A \times \Theta)$ such that $\nu^k \in SI(\mu^k)$, we have $\sum_{\theta\in\Theta} \nu^k(\mathbf{0}, \theta) \to 1$ as $k \to \infty$.[18] Then, by definition (along with a

[17]Precisely, Oyama and Takahashi (2020) assumed (strict) sequential obedience and constructed such information structures with an additional property that players know their own payoffs to comply with the definition of robustness of Kajii and Morris (1997).

[18]This definition of robustness differs from the original one in Kajii and Morris (1997) in a few respects. In defining "nearby" incomplete information games, (i) there is no restriction on payoffs other than those at θ^* in Kajii and Morris (1997), while payoffs are restricted to supermodular payoffs that are fixed by finitely many payoff states (including a dominance state) in our definition; and (ii) information structures are restricted to those such that players know that their payoffs are given by the complete information game with high probability in Kajii and Morris (1997), while this "known own payoffs" assumption is not imposed in ours. For BAS games, the difference in (i) is immaterial, while (ii) makes our definition stronger than that of Kajii and Morris (1997), but only in nongeneric cases. In particular, our definition does not depend on the choice of the state space Θ and the base game $(d_i)_{i\in i}$ embedding the given complete information game $(f_i)_{i\in I}$ as long as Θ is finite and the dominance state assumption is satisfied.

compactness argument), $\mathbf{0}$ is not robust in $(f_i)_{i\in I}$ if and only if there exists an outcome $\xi \in \Delta(A)$ with $\xi(\mathbf{0}) < 1$ that is limit S-implementable at θ^*. Thus, by the "only if" part of Corollary 4(1), $\mathbf{0}$ is not robust in $(f_i)_{i\in I}$ only if there exists an ordered outcome $\rho \in \Delta(\Gamma)$ with $\rho(\emptyset) < 1$ that satisfies weak sequential obedience in $(f_i)_{i\in I}$, and vice versa by Corollary 4(2). But when the latter condition holds, the ordered outcome $\rho' \in \Delta(\Gamma)$ given by $\rho'(\emptyset) = 0$ and $\rho'(\gamma) = \frac{\rho(\gamma)}{1-\rho(\emptyset)}$ for $\gamma \neq \emptyset$ also satisfies weak sequential obedience in $(f_i)_{i\in I}$. Therefore, we have the following.

Corollary 5. *Action profile* $\mathbf{0}$ *is not robust in* $(f_i)_{i\in I}$ *if and only if there exists some outcome* $\xi \in \Delta(A)$ *with* $\xi(\mathbf{0}) = 0$ *that satisfies weak sequential obedience in* $(f_i)_{i\in I}$.

In the two-player case, $\mathbf{0}$ is the unique outcome that satisfies weak sequential obedience if and only if it is risk dominant. Thus, the "only if" part of Corollary 5 implies the robustness of a risk dominant equilibrium (Kajii and Morris (1997)).

Oyama and Takahashi (2020) observed via a duality theorem that the latter condition in Corollary 5 holds if and only if for any $(\lambda_i)_{i\in I} \in \mathbb{R}^I_+$,

$$\max_{a\neq\mathbf{0}} \max_{\gamma:a(\gamma)=a} \sum_{i\in S(a)} \lambda_i f_i(a_{-i}(\gamma)) \geq 0. \tag{3.7}$$

This condition is closely related to the concept of "monotone potential", introduced by Morris and Ui (2005) as a sufficient condition for robustness for supermodular games with any (finite) number of actions.[19] Indeed, it is equivalent to the condition that $\mathbf{0}$ is *not* a monotone potential maximizer in the complete information game $(f_i)_{i\in I}$.[20] Thus, the sufficiency (Morris and Ui (2005)) and necessity (Oyama and Takahashi (2020)) of monotone potential maximization for robustness are in fact obtained as a corollary to our Theorem 1 (modulo the difference in the definition of robustness).

Corollary 6. *Action profile* $\mathbf{0}$ *is robust in* $(f_i)_{i\in I}$ *if and only if it is a monotone potential maximizer in* $(f_i)_{i\in I}$.

In light of this, our S-implementability question can be viewed as an incomplete information generalization of the robustness question.[21]

[19]Morris and Ui (2005) build on Ui (2001), who established the robustness of potential maximizing equilibria in (possibly non-supermodular) potential games (with many actions).

[20]Here we refer to the strict version of the monotone potential condition employed by Oyama and Takahashi (2020). Formally, $\mathbf{0}$ is a *monotone potential maximizer* in $(f_i)_{i\in I}$ if there exist a function $\phi\colon A \to \mathbb{R}$ (*monotone potential*) and $(\lambda_i)_{i\in I} \in \mathbb{R}^I_+$ such that $\lambda_i f_i(a_{-i}) \leq \phi(1, a_{-i}) - \phi(0, a_{-i})$ for all i and a_{-i}, and $\phi(\mathbf{0}) > \phi(a)$ for all $a \neq \mathbf{0}$.

[21]Morris and Ui (2020) propose and study an incomplete information generalization of the robustness notion of Kajii and Morris (1997), which neither nests nor is nested by our exercise.

3.3.3. *Joint Design of Information and Payoffs.* Our approach has been to fix the base game payoffs $(d_i)_{i \in I}$ and has shown that sequential obedience characterizes the set of outcomes that are S-implementable. Alternatively, if a target outcome is fixed, the sequential obedience condition can read as characterizing the set of payoffs under which the target outcome is S-implementable. In fact, the recent work by Halac et al. (2021) and Moriya and Yamashita (2020) on the optimal joint design of transfers and information in the context of team production can be understood from this viewpoint. In particular, the results of Halac et al. (2021), who allow for private contracts in the complete information setting of Winter (2004), can be derived in terms of limit S- (or full) implementation and hence sequential obedience in complete information games. Suppose that, as in Winter (2004), the principal aims at S- (hence fully) implementing action profile $\mathbf{1}$ ("all agents exerting effort") in the complete information game determined by a smallest amount of bonus payments. But suppose that the principal can also offer a large bonus to each agent privately with positive but vanishingly small probability so that exerting effort is a dominant action, while the principal can commit to any information structure that determines the agents' beliefs about the bonus payments offered to others. Then the optimization problem reduces to minimizing the total bonus payment subject to the constraint that $\mathbf{1}$ is limit S-implementable along the sequence of induced incomplete information games, and hence by Corollary 4, satisfies the sequential obedience condition (3.6) in the limit complete information game. Relative to the case of public contracts, allowing private contracts weakens the deterministic divide-and-conquer condition (3.5) to the stochastic sequential obedience condition (3.6), and in particular, when the underlying game is symmetric, it makes the principal strictly better off by offering symmetric bonuses, over the divide-and-conquer scheme which discriminates ex ante symmetric agents. In Morris et al. (2022b), we formally describe the above solution method. By appealing also to the concept of potential to be introduced in Section 3.5, we obtain additional insights over Halac et al. (2021).[22]

3.4. **A Dual Representation of Sequential Obedience.** We now report a dual representation of sequential obedience. Sequential obedience of an outcome ν is defined by the existence of an ordered outcome ν_Γ inducing ν that satisfies condition (3.1), or in

[22] In Morris et al. (2022b), we also discuss the incomplete information generalization of Winter (2004) by Moriya and Yamashita (2020) and show that a straightforward application of our results immediately solves their model with an extension with many players and many states.

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 3.5, where we provide a simpler characterization of sequential obedience when the base game has a potential. It also allows us to further highlight the important connection with Oyama and Takahashi (2020).

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 ν 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 $i \in I(\nu)$,*

$$\sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \max_{\gamma : a(\gamma) = a} \sum_{i \in S(a)} \lambda_i d_i(a_{-i}(\gamma), \theta) > (\textit{resp. } \geq)\ 0. \tag{3.8}$$

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. For given $(\lambda_1, \lambda_2) \in \mathbb{R}^2_+ \setminus \{(0, 0)\}$, the left hand side of (3.8) is computed as

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

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 outcome (2.2) and satisfies sequential obedience (resp. weak sequential obedience) if $\delta > 0$ (resp. if $\delta = 0$).

As discussed in the previous Section 3.3.2, for the limit complete information case where $\mu(\theta^*) = 1$, Oyama and Takahashi (2020) introduced a dual characterization of sequential obedience in complete information games. To understand the connection, observe that from Proposition 2 it follows that there exists some $\xi \in \Delta(A)$ with $\xi(\mathbf{0}) = 0$ that satisfies weak sequential obedience in $(f_i)_{i\in I}$ if and only if

$$\max_{\xi\in\Delta(A\backslash\{\mathbf{0}\})}\min_{\lambda\in\mathbb{R}^I_+} F(\xi,\lambda)\geq 0, \tag{3.9}$$

where

$$F(\xi,\lambda)=\sum_{a\neq\mathbf{0}}\xi(a)\max_{\gamma:a(\gamma)=a}\sum_{i\in S(a)}\lambda_i f_i(a_{-i}(\gamma)),$$

while Oyama and Takahashi's (2020) characterization (3.7) is equivalently written as

$$\min_{\lambda\in\mathbb{R}^I_+}\max_{\xi\in\Delta(A\backslash\{\mathbf{0}\})} F(\xi,\lambda)\geq 0. \tag{3.10}$$

Since $F(\xi,\lambda)$ is linear in ξ and convex in λ, the expressions (3.9) and (3.10) are equivalent by the minimax theorem.

3.5. **Sequential Obedience in Potential Games.** Our sequential obedience characterization of S-implementability applies to all BAS games. Here we will report a simpler version of sequential obedience for BAS potential games. A potential game has the property that the sum of payoff gains for a sequence of players switching from 0 to 1 is independent of the order in which players switch. This will allow us to provide a characterization of sequential obedience in terms of the change in the potential by a simultaneous switch of a subset of players.

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 identify a potential game with its potential function Φ. We adopt the normalization that $\Phi(\mathbf{0},\theta)=0$ for all $\theta\in\Theta$.

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 $a\in A$ and (abusing notation slightly) $n(a_{-i})$ for the number of players choosing action 1 in action profile $a_{-i}\in A_{-i}$.

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 θ. Assume that $R(|\Theta|)+h_1>c_i$ for all $i\in I$, so that the dominance state assumption holds with $\overline{\theta}=|\Theta|$. 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 $R(\mathbf{b}) = 0$, $R(\mathbf{g}) = 9$, $h_1 = 0$, $h_2 = 3$, $c_1 = 7$, and $c_2 = 8$. It has a potential

b	Not	Invest
Not	0	−8
Invest	−7	−12

g	Not	Invest
Not	0	1
Invest	2	6

. (3.11)

Example 2 (Regime Change Game)**.** Let $\Theta = \{1, \dots, |\Theta|\}$, and

$$d_i(a_{-i}, \theta) = \begin{cases} c_i & \text{if } n(a_{-i}) + 1 > |I| - k(\theta), \\ c_i - 1 & \text{if } n(a_{-i}) + 1 \leq |I| - k(\theta), \end{cases}$$

where $0 < c_i < 1$, and $k \colon \Theta \to \mathbb{N}$ is strictly increasing. We assume that $k(1) \geq 1$ and $k(|\Theta|) = |I|$, so that the dominance state assumption holds with $\overline{\theta} = |\Theta|$. 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 larger than or equal to $k(\theta)$, or equivalently, the number of non-attackers (action 1 players) is smaller than or equal to $|I| - k(\theta)$. Attack yields a gross benefit 1 (resp. 0) upon regime change (resp. status quo) with cost c_i, while the payoff of abstention is always 0. This game has a potential:

$$\Phi(a, \theta) = \begin{cases} \sum_{i \in S(a)} c_i - (|I| - k(\theta)) & \text{if } n(a) > |I| - k(\theta), \\ \sum_{i \in S(a)} c_i - n(a) & \text{if } n(a) \leq |I| - k(\theta). \end{cases}$$

This is a finite-state, finite-player version of the continuous-state, continuum-player regime change game studied by Morris and Shin (1998, 2004) and analyzed in the context of information design by Inostroza and Pavan (2022) and Li et al. (2022).

We now provide a simpler characterization of sequential obedience for potential games. For any outcome $\nu \in \Delta(A \times \Theta)$, define

$$\Phi_\nu(a) = \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta)\Phi(a \wedge a', \theta),$$

where $b = a \wedge a'$ denotes the action profile such that $b_i = 1$ if and only if $a_i = a'_i = 1$. To interpret this function Φ_ν, imagine a hypothetical situation where players make commitments whether to "play $a_i = 1$ whenever recommended to do so" (represented simply as $a_i = 1$) or "play $a_i = 0$ whatever the recommendation" (represented as $a_i = 0$), before they receive recommendations a' according to ν. Thus, if the profile of commitments is a, then the ex post play is $a \wedge a'$ when the profile of recommendations is a', and hence the ex ante expected value of the potential Φ with respect to ν is $\Phi_\nu(a)$. In particular, $\Phi_\nu(\mathbf{1})$ is the expected potential of ν when all players follow the recommendations, while $\Phi_\nu(0, \mathbf{1}_{-i})$ is that when only player i deviates to action 0 with all others following recommendations; therefore, upper obedience—the requirement that players have an incentive to follow recommendation of action 1—can be written with function Φ_ν as

$$\Phi_\nu(\mathbf{1}) \geq \Phi_\nu(0, \mathbf{1}_{-i})$$

for all $i \in I$. Sequential obedience is shown to be equivalent to the stronger condition that the outcome potential is maximized when all players follow the recommendations (recall that $I(\nu)$ is the set of players who are recommended to play action 1 with positive probability under ν).

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

$$\Phi_\nu(\mathbf{1}) > (\textit{resp. } \geq)\ \Phi_\nu(a) \tag{3.12}$$

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

The proof is given in Appendix A.3, where we verify that condition (3.12) 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 Φ_ν.

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

	Not	Invest
Not	0	$-\frac{3}{2} + 8\delta$
Invest	$-\frac{3}{4} + 7\delta$	12δ

.

Thus, outcome ν satisfies condition (3.12) with strict (resp. weak) inequality if $\delta > 0$ (resp. if $\delta = 0$).

In the special case where there is limit complete information at some θ^* (discussed in Section 3.3) and the outcome is (the degenerate outcome on) pure action profile $\mathbf{1}$, condition (3.12) reduces to the condition that

$$\Phi(\mathbf{1}, \theta^*) > \text{(resp. } \geq \text{)} \ \Phi(a, \theta^*)$$

for all $a \neq \mathbf{1}$, that is, that $\mathbf{1}$ is potential maximizing in the complete information potential game $\Phi(\cdot, \theta^*)$. Thus, by Corollary 4 and Proposition 3, $\mathbf{1}$ is limit S-implementable at θ^* if and only if it is a weak potential maximizer at θ^*.[23]

3.6. **Sequential Obedience in Convex Potential Games.** The sequential obedience condition can be further simplified if the potential satisfies a convexity condition that bounds the degree of asymmetry in the game and if the outcome is a perfect coordination outcome. We first discuss and define these properties.

Our convexity condition requires that for all $\theta \in \Theta$, $\Phi(a, \theta)$ be smaller than a convex combination of $\Phi(\mathbf{0}, \theta) = 0$ and $\Phi(\mathbf{1}, \theta)$.[24]

Definition 8. Potential Φ satisfies *convexity* if

$$\Phi(a, \theta) \leq \frac{n(a)}{|I|} \Phi(\mathbf{1}, \theta) \tag{3.13}$$

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 $\Phi(a, \theta) = \widehat{\Phi}(n(a), \theta)$ for some function $\widehat{\Phi} \colon \{0, \ldots, |I|\} \times \Theta \to \mathbb{R}$, then supermodularity implies that $\widehat{\Phi}(n+1, \theta) - \widehat{\Phi}(n, \theta)$ is increasing in n and thus (3.13) is satisfied. If payoffs are asymmetric, define a symmetrized potential $\widehat{\Phi} \colon \{0, \ldots, |I|\} \times \Theta \to \mathbb{R}$ by

$$\widehat{\Phi}(n, \theta) = \frac{1}{\binom{|I|}{n}} \sum_{a : n(a) = n} \Phi(a, \theta).$$

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

$$\Delta(a, \theta) = \Phi(a, \theta) - \widehat{\Phi}(n(a), \theta).$$

[23] In Morris et al. (2022b), we report interesting connections between the sequential obedience condition in complete information potential games and some well-known concepts from cooperative game theory, in particular the *core* of the supermodular set function (hence cooperative game) $S \mapsto \Phi(\mathbf{1}_S, \theta^*)$ (where for $S \subset I$, $\mathbf{1}_S$ denotes the action profile a such that $a_i = 1$ if and only if $i \in S$).

[24] This condition is thus a strengthening of the requirement that $\arg\max_{a \in A} \Phi(a, \theta) \cap \{\mathbf{0}, \mathbf{1}\} \neq \emptyset$ 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), Leister et al. (2022)).

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

$$M(n,\theta) = \frac{n}{|I|}\Phi(\mathbf{1},\theta) - \widehat{\Phi}(n,\theta) \geq 0$$

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

$$\Phi(a,\theta) = \Delta(a,\theta) + \widehat{\Phi}(n(a),\theta) \leq \frac{n(a)}{|I|}\Phi(\mathbf{1},\theta)$$

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

$$\frac{1}{\ell}\sum_{k=1}^{\ell}(h_k - c_k) \leq \frac{1}{|I|}\sum_{k=1}^{|I|}(h_k - c_k) \tag{3.14}$$

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. For the investment game example (2.1) in Section 2, the potential as given in (3.11) can be verified to satisfy convexity by the definition, or by this sufficient condition, $h_1 - c_1\ (= -7) < h_2 - c_2\ (= -5)$.

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

An outcome is perfectly coordinated if either all play 0 or all play 1.

Definition 9. Outcome ν satisfies *perfect coordination* if for all $\theta \in \Theta$, $\nu(a,\theta) > 0$ only for $a \in \{\mathbf{0}, \mathbf{1}\}$.

As we noted, this property has been introduced in the context of regime change games by Inostroza and Pavan (2022). Now we have:

Proposition 4. *In a convex potential game, a perfectly coordinated outcome ν satisfies sequential obedience (resp. weak sequential obedience) if and only if it satisfies condition* (3.12) *in Proposition 3 for $a = \mathbf{0}$, that is, the average potential of $\mathbf{1}$ under ν is positive (resp. nonnegative):*

$$\Phi_\nu(\mathbf{1}) = \sum_{\theta \in \Theta} \nu(\mathbf{1}, \theta)\Phi(\mathbf{1}, \theta) > (\textit{resp.} \geq)\ 0. \tag{3.15}$$

The proof is given in Appendix A.4.

4. Information Design with Adversarial Equilibrium Selection

In this section, we study the optimal information design problem with adversarial equilibrium selection as set out in Section 1.4. Our implementation results allow us to express the problem as a finite dimensional linear problem (Corollary 3). In the following, we adopt the normalization that $V(\mathbf{0}, \theta) = 0$ for all $\theta \in \Theta$, and for simplicity, assume that $V(\mathbf{1}, \theta) > 0$ for all $\theta \in \Theta$.

Assume that the base game has a potential Φ. Our main characterization of optimal outcomes requires one additional assumption on the designer's objective V:

Definition 10. Designer's objective V satisfies *restricted convexity* with respect to potential Φ if

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

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

Convexity of V, $V(a, \theta) \leq \frac{n(a)}{|I|} V(\mathbf{1}, \theta)$ for all a and θ, is obviously a sufficient condition for restricted convexity, irrespective of Φ. As discussed above when discussing the convexity of Φ, 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 $\alpha \geq 1$. This includes both the case where the designer wants to maximize the expected fraction of players who play action 1 ($\alpha = 1$), and thus has no preference over whether the players are coordinated or not; and the case where the designer cares only about 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 maintaining the status quo.[25]

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

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

Now assume that the potential Φ satisfies convexity and the objective V satisfies restricted convexity with respect to Φ. Under the convexity of Φ, coordinating players' actions tends to slacken the incentive constraints, and by the restricted convexity of V, it also improves the value for the designer. Indeed, as to be shown in Theorem 2, there will be 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.

Consider the maximization problem with respect to perfectly coordinated outcomes subject to consistency and weak sequential obedience (in the form of condition (3.15) in Proposition 4):

$$\max_{(\nu(\mathbf{1},\theta))_{\theta\in\Theta}} \sum_{\theta\in\Theta} \nu(\mathbf{1}, \theta) V(\mathbf{1}, \theta) \tag{4.1a}$$

subject to

$$0 \leq \nu(\mathbf{1}, \theta) \leq \mu(\theta) \quad (\theta \in \Theta), \tag{4.1b}$$

$$\sum_{\theta\in\Theta} \nu(\mathbf{1}, \theta) \Phi(\mathbf{1}, \theta) \geq 0. \tag{4.1c}$$

Notice that the problem can be viewed as a Bayesian persuasion problem where the role of the receiver is played by the potential and there are two available actions, $\mathbf{0}$ and $\mathbf{1}$. The solution will clearly have $\nu(\mathbf{1}, \theta) = \mu(\theta)$ for all "good states" θ with $\Phi(\mathbf{1}, \theta) \geq 0$ and as many "bad states" θ with $\Phi(\mathbf{1}, \theta) < 0$ as possible consistent with the average potential $\sum_{\theta\in\Theta} \nu(\mathbf{1}, \theta) \Phi(\mathbf{1}, \theta)$ being nonnegative. But which bad states to include? We will see that it is optimal to include states with the lowest cost-benefit ratio, where the cost is $-\Phi(\mathbf{1}, \theta)$ and the benefit is $V(\mathbf{1}, \theta)$.

[25]This objective is studied in the regime change applications of Inostroza and Pavan (2022) and Li et al. (2022) (Inostroza and Pavan (2022) also consider some more general objectives).

Concretely, we relabel the states as $\Theta = \{1, \dots, |\Theta|\}$ in such a way that $\frac{\Phi(\mathbf{1},\theta)}{V(\mathbf{1},\theta)}$ is increasing in θ:[26]

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

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

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

If $\Psi(0) = \sum_{\theta' \in \Theta} \mu(\theta')\Phi(\mathbf{1},\theta') \geq 0$, then the outcome "always play 1" is an optimal solution. In the following, we assume that $\Psi(0) < 0$. Let $\theta^* \in \Theta$ be the unique state such that $\Psi(\theta) \geq 0$ if and only if $\theta \geq \theta^*$. Note that $\Phi(\mathbf{1},\theta^*) < 0$. Let

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

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

Now define the perfectly coordinated outcome ν^* by

$$\nu^*(a,\theta) = \begin{cases} \mu(\theta) & \text{if } a = \mathbf{1} \text{ and } \theta > \theta^*, \\ p^* & \text{if } a = \mathbf{1} \text{ and } \theta = \theta^*, \\ \mu(\theta) - p^* & \text{if } a = \mathbf{0} \text{ and } \theta = \theta^*, \\ \mu(\theta) & \text{if } a = \mathbf{0} \text{ and } \theta < \theta^*, \\ 0 & \text{otherwise,} \end{cases} \tag{4.2}$$

which clearly satisfies consistency (4.1b). This outcome satisfies the weak sequential obedience constraint (4.1c) with equality:

$$\sum_{\theta \in \Theta} \nu^*(\mathbf{1},\theta)\Phi(\mathbf{1},\theta) = \Psi(\theta^*) + p^*\Phi(\mathbf{1},\theta^*) = 0. \tag{4.3}$$

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

$$\sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu^*((0,a_{-i}),\theta)d_i(a_{-i},\theta) = \sum_{\theta \leq \theta^*} \nu^*(\mathbf{0},\theta)\Phi((1,\mathbf{0}_{-i}),\theta) < 0,$$

since by the convexity of Φ, $\Phi((1,\mathbf{0}_{-i}),\theta) \leq \frac{1}{|I|}\Phi(\mathbf{1},\theta) < 0$ for all $\theta \leq \theta^*$. Thus, $\nu^* \in \overline{SI}$ by Proposition 4 and Corollary 1. Theorem 2 shows that ν^* is an optimal solution to the problem (4.1) and that it is an optimal outcome of the adversarial information design problem.

[26] As is clear from the argument below, the choice of the order on the states θ for which $\Phi(\mathbf{1},\theta) \geq 0$ is inconsequential.

Theorem 2. *Consider a game with convex potential* Φ *and a designer with objective* V *satisfying restricted convexity with respect to* Φ*. Then there exists an optimal outcome of the adversarial information design problem that satisfies perfect coordination. In particular, the outcome* ν^* *defined in* (4.2) *is an optimal outcome.*

The proof is given in Appendix A.5.

We can illustrate the result with the two-player two-state 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 Φ given in (3.11) in Example 1 (which satisfies convexity as verified in Example 3), we have $\Psi(\mathbf{b}) = -12$ and $\Psi(\mathbf{g}) = 6$, and hence $\theta^* = \mathbf{b}$ is the threshold state. With $p^* = \frac{1}{4}$, the optimal outcome ν^* is thus as given in (2.3) in Section 2.

To compare Theorem 2 with the existing literature, it is useful to consider a continuous version of our problem with a continuum of symmetric players and a continuous state space $\Theta \subset \mathbb{R}$ where the (common) payoff difference function d and the designer's objective V depend on the proportion ℓ of players playing action 1 and are nondecreasing in the state θ, assuming also that V satisfies restricted convexity. This case, in particular, encompasses the regime change game studied by Li et al. (2022). In this case, the potential is written as $\Phi(\ell, \theta) = \int_0^\ell d(\ell', \theta) d\ell'$, which is convex in ℓ. The continuous limit of the optimal outcome as characterized in Theorem 2 then becomes the outcome that has all players playing action 1 (resp. 0) if the state is above (resp. below) the threshold state θ^* that solves

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

(with μ denoting the probability distribution of θ also in this case).[27] In the special case of the regime change game, this is the outcome identified by Li et al. (2022), who present an alternative implementation of the same outcome which is tailored to the class of games they study. They also show that a finite signal version of their construction is a unique optimal information structure when the designer is constrained to only use finite information structures with up to K signals, and their unconstrained optimal information structure is characterized as the limit of the optimal finite information structures as $K \to \infty$. In a companion note, Morris et al. (2022a), we provide a simple, global game implementation of the optimal outcome which applies to all symmetric and state-monotonic BAS games.

[27] In Morris et al. (2022a), we give a formal derivation of this result.

5. Discussion

In this section, we discuss how our results would generalize or vary under alternative assumptions and formulations. Formal treatments of those issues are relegated to the Supplemental Appendix.

5.1. **Full Implementation.** In our analysis, we focused on S-implementation, rather than full implementation. It is the relevant notion, in particular, when an information designer is concerned with inducing the high action in the worst case scenario. But we show in Supplemental Appendix B.3 that the arguments for full implementation are straightforward extensions of the results for S-implementation. An outcome is fully implementable if it satisfies not only sequential obedience which is necessary for S-implementation, but also the reverse version of sequential obedience which is necessary for "largest equilibrium implementation", and conversely, under an appropriate extension of the assumptions of dominance state and grain of dominance, these necessary conditions are also jointly sufficient for full implementation. We further show that for any outcome in $\overline{SI}$, there exists an outcome in $\overline{FI}$ that stochastically dominates that outcome. Thus, under the action monotonicity of the objective function, optimal information design subject to S-implementability is in fact equivalent to that subject to full implementability.

5.2. **Non-Supermodular Payoffs.** The supermodularity of payoffs has been maintained throughout the paper. For a general binary-action game $(d_i)_{i \in I}$ with possibly non-supermodular payoffs, our arguments still continue to work in the special case where we are interested in implementing the "all players always play action 1" outcome by a unique rationalizable strategy profile. (Note that under the supermodularity assumption, this is equivalent to S-implementation and full implementation.) In this case, the implementability is characterized by the strengthening of sequential obedience that requires that action 1 be a strict best response for a player whenever action 1 is played by others before him in the sequence, but independent of the play of players after him.[28] Then applying our results to the BAS game $(\underline{d}_i)_{i \in I}$ obtained by $\underline{d}_i(a_{-i}, \theta) = \min_{a'_{-i} \geq a_{-i}} d_i(a'_{-i}, \theta)$ will give the characterization; see Supplemental Appendix B.4.1 for formal arguments.

5.3. **Many Actions.** For games with more than two actions, there are several possible generalizations of sequential obedience that will be necessary for S-implementability. A

[28] A similar condition appears in Halac et al. (2021, Section V) and Halac et al. (2022).

generalized sequential obedience condition would be to require the existence of a distribution over sequences of action profiles, possibly correlated with the state, such that each player whenever recommended to switch from an action to a higher action has a strict incentive to do so when expecting that the switches before the current switch have occurred; see Supplemental Appendix B.4.2 for a formal account. Then, the necessity of this condition for S-implementability can be proved almost identically as in the proof of Theorem 1(1): consider the sequential best response process from the smallest strategy profile, and then averaging the obedience conditions upon switches leads to the generalized sequential obedience condition.

On the other hand, establishing the sufficiency of generalized sequential obedience (along with consistency, obedience, and appropriately modified versions of dominance state and grain of dominance) is not obvious. A proof strategy of the same approach as in the proof of Theorem 1(2) would be to consider an information structure generated by multi-dimensional signals, with each dimension suggesting the timing of switching from an action to another in the random sequence of action profiles. However, this would not work as desired in general, since the averaged condition of sequential obedience may well be too coarse to control the incentives there. In Supplemental Appendix B.4.2, we report a special case which in effect reduces to a binary-action case, but still covers the result of Hoshino (2022). We have to leave for future research identifying a broader class of games in which our current approach works (with minimal modifications), or developing a new idea in constructing information structures, possibly along with a more refined sequential obedience-like condition.[29]

5.4. **Adversarial Information Sharing.** Our implicit assumption has been that players do not share among themselves the information that is privately revealed to them by the information structure. Given that the designer is concerned with the worst case in the actions of players, it would be possible that she has a robustness concern also about the possibility of information sharing among the players. A simple way to allow for this possibility is to suppose that there might be a non-strategic information sharing protocol that can selectively reveal players' information to others and ask what is the designer's optimal choice of information structure when she assumes that there will be adversarial

[29]One might appeal to the approach of Gossner and Veiel (2022), who developed a finite automaton representation of "critical" information structures that characterize rationalizable outcomes in general finite games.

information sharing.[30] In Supplemental Appendix B.4.3, we formulate this problem and prove that, under the supermodularity of the payoffs and the monotonicity of the objective function, the designer cannot do better than revealing a public signal to the players in this case. The problem then reduces to a Bayesian persuasion problem.

5.5. **Finite Information Structures.** Our implementation in the proof of Theorem 1(2) involves infinitely many types, but it is straightforward to construct its finite version, depending on the environment and the outcome to be implemented. The crucial assumption is that there is no a priori bound on the number of types; see Supplemental Appendix B.4.4 for a formal argument. Specialized to a symmetric two-player two-state example, Mathevet et al. (2020) present an information structure implementing the optimal outcome that is much smaller than the one that the finite version of our construction would give. It is an interesting open problem to characterize the smallest number of types needed to implement a given outcome for general BAS games.

Appendix

A.1. **Proof of Theorem 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 $\Delta(A_i)$).

A.1.1. *Proof of Theorem 1(1).* Let $\nu \in \Delta(A \times \Theta)$ be S-implementable, and let (T, π) be a type space whose smallest equilibrium $\underline{\sigma}$ induces ν. By Proposition 1, ν 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 $\sigma_i^0(t_i) = 0$ 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 σ_{-i}^{n-1}. Thus,

$$\sigma_i^n(t_i) = \begin{cases} 1 & \text{if } i \equiv n \pmod{|I|}, \\ & \text{and } \sum_{t_{-i},\theta} \pi((t_i, t_{-i}), \theta) d_i(\sigma_{-i}^{n-1}(t_{-i}), \theta) > 0, \\ \sigma_i^{n-1}(t_i) & \text{otherwise.} \end{cases}$$

By supermodularity, for each i and t_i, the sequence $\{\sigma_i^n(t_i)\}$ (of pure actions 0 and 1) is monotone increasing and converges to $\underline{\sigma}_i(t_i)$. Let $n_i(t_i) = n$ if $\sigma_i^{n-1}(t_i) = 0$ and $\sigma_i^n(t_i) = 1$

[30]Galperti and Perego (2020) study a non-strategic model where information is shared automatically among players through a fixed network which is known to the designer, and Mathevet and Taneva (2022) consider a similar model but account for incentives. Both papers consider partial implementation.

(and hence $\underline{\sigma}_i(t_i) = 1$); let $n_i(t_i) = \infty$ if $\sigma_i^n(t_i) = 0$ for all n (and hence $\underline{\sigma}_i(t_i) = 0$). Write $n(t) = (n_1(t_1), \dots, n_{|I|}(t_{|I|}))$. For $\gamma = (i_1, \dots, i_k) \in \Gamma$, let $T(\gamma)$ denote the set of type profiles t such that $n(t)$ is ordered according to γ: i.e., those type profiles t such that $n_i(t_i) = \infty$ for all $i \notin \{i_1, \dots, i_k\}$, and $n_{i_\ell}(t_{i_\ell}) < n_{i_m}(t_{i_m}) < \infty$ if and only if $\ell < m$.

Now, define $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ by

$$\nu_\Gamma(\gamma, \theta) = \sum_{t \in T(\gamma)} \pi(t, \theta)$$

for each $(\gamma, \theta) \in \Gamma \times \Theta$. Observe that ν_Γ induces ν: indeed, for each $(a, \theta) \in A \times \Theta$, we have

$$\begin{aligned}\sum_{\gamma: a(\gamma)=a} \nu_\Gamma(\gamma, \theta) &= \sum_{\gamma: a(\gamma)=a} \sum_{t \in T(\gamma)} \pi(t, \theta) \\ &= \sum_{t: n_i(t_i) < \infty \iff a_i = 1} \pi(t, \theta) = \sum_{t: \underline{\sigma}(t)=a} \pi(t, \theta) = \nu(a, \theta).\end{aligned}$$

To show sequential obedience, fix any $i \in I$ with $\nu_\Gamma(\Gamma_i \times \Theta) > 0$. Note that for all $t_i \in T_i$ with $n_i(t_i) < \infty$, we have

$$\sum_{t_{-i} \in T_{-i}, \theta \in \Theta} \pi((t_i, t_{-i}), \theta) d_i\left(\sigma_{-i}^{n_i(t_i)-1}(t_{-i}), \theta\right) > 0.$$

By adding up the inequality over all such t_i, we have

$$\begin{aligned}0 &< \sum_{t_i: n_i(t_i) < \infty} \sum_{t_{-i} \in T_{-i}, \theta \in \Theta} \pi((t_i, t_{-i}), \theta) d_i\left(\sigma_{-i}^{n_i(t_i)-1}(t_{-i}), \theta\right) \\ &= \sum_{\gamma \in \Gamma_i} \sum_{t \in T(\gamma)} \sum_{\theta \in \Theta} \pi(t, \theta) d_i(a_{-i}(\gamma), \theta) \\ &= \sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta).\end{aligned}$$

Thus, ν 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 $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ be an ordered outcome establishing sequential obedience. Since $\nu(\mathbf{1}, \overline{\theta}) > 0$ by grain of dominance, there exists $\bar{\gamma} \in \Gamma$ containing all players with $\nu_\Gamma(\bar{\gamma}, \overline{\theta}) > 0$. For $\varepsilon > 0$ with $\varepsilon < \nu_\Gamma(\bar{\gamma}, \overline{\theta})$, let

$$\tilde{\nu}_\Gamma(\gamma, \theta) = \begin{cases} \dfrac{\nu_\Gamma(\gamma, \theta) - \varepsilon}{1 - \varepsilon} & \text{if } (\gamma, \theta) = (\bar{\gamma}, \overline{\theta}), \\ \dfrac{\nu_\Gamma(\gamma, \theta)}{1 - \varepsilon} & \text{otherwise,} \end{cases}$$

where we assume that ε is sufficiently small that $\tilde{\nu}_\Gamma$ satisfies sequential obedience, i.e.,

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

for all $i \in I$. By the dominance state assumption, we can take a $\bar{q} < 1$ such that

$$\bar{q} d_i(\mathbf{0}_{-i}, \overline{\theta}) + (1 - \bar{q}) \min_{\theta \neq \overline{\theta}} d_i(\mathbf{0}_{-i}, \theta) > 0 \tag{A.1}$$

for all $i \in I$. Then let $\eta > 0$ be such that

$$\frac{\frac{\varepsilon}{|I|-1}}{\frac{\varepsilon}{|I|-1} + \eta} \geq \bar{q} \tag{A.2}$$

and

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1 - \eta)^{|I| - n(a_{-i}(\gamma)) - 1} \tilde{\nu}_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > 0 \tag{A.3}$$

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, π) as follows. For each $i \in I$, let

$$T_i = \begin{cases} \{1, 2, \ldots\} & \text{if } \tilde{\nu}_\Gamma(\Gamma_i \times \Theta) = 1, \\ \{1, 2, \ldots\} \cup \{\infty\} & \text{otherwise.} \end{cases}$$

Let $\pi \in \Delta(T \times \Theta)$ be given by

$$\pi(t, \theta) = \begin{cases} (1 - \varepsilon)\eta(1 - \eta)^m \tilde{\nu}_\Gamma(\gamma, \theta) & \text{if there exist } m \in \mathbb{N} \text{ and } \gamma \in \Gamma \setminus \{\emptyset\} \\ & \text{such that } t_i = m + \ell(i, \gamma) \text{ for all } i \in I, \\ (1 - \varepsilon)\tilde{\nu}_\Gamma(\emptyset, \theta) & \text{if } t_1 = \cdots = t_{|I|} = \infty, \\ \dfrac{\varepsilon}{|I| - 1} & \text{if } 1 \leq t_1 = \cdots = t_{|I|} \leq |I| - 1 \text{ and } \theta = \overline{\theta}, \\ 0 & \text{otherwise} \end{cases}$$

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 π is consistent with μ: $\sum_t \pi(t, \theta) = \sum_\gamma \nu_\Gamma(\gamma, \theta) = \mu(\theta)$ for all $\theta \in \Theta$.

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

Proof. For $\tau \in \{1, \ldots, |I| - 1\}$, we have

$$\pi(\overline{\theta} | t_i = \tau) = \frac{\sum_{t_{-i}} \pi(t_i = \tau, t_{-i}, \overline{\theta})}{\sum_{t_{-i}, \theta} \pi(t_i = \tau, t_{-i}, \theta)} \geq \frac{\frac{\varepsilon}{|I|-1}}{\frac{\varepsilon}{|I|-1} + \eta} \geq \bar{q},$$

where the first inequality holds since $\sum_{t_{-i}} \pi(t_i = \tau, t_{-i}, \overline{\theta}) \geq \pi(t_1 = \cdots = t_{|I|} = \tau, \overline{\theta}) = \frac{\varepsilon}{|I|-1}$, and $\sum_{t_{-i}, \theta} \pi(t_i = \tau, t_{-i}, \theta) \leq \frac{\varepsilon}{|I|-1} + (1 - \varepsilon)\eta(1 - \eta)^{\tau - 1} \sum_{\gamma \in \Gamma_i, \theta} \tilde{\nu}_\Gamma(\gamma, \theta) \leq \frac{\varepsilon}{|I|-1} + \eta$, while the second inequality is by (A.2). □

For $S \subset I$, we denote by $\mathbf{1}_S$ 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 $\tau \in \{|I|, |I|+1, \ldots\}$,*

$$\pi(\{j \neq i \mid t_j < \tau\} = S, \theta | t_i = \tau) = (1-\eta)^{|I|-|S|-1} \tilde{\nu}_\Gamma(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = \mathbf{1}_S\} \times \{\theta\})/C_i$$

for all $S \subset I \setminus \{i\}$, where $C_i = \sum_{\ell=1}^{|I|} (1-\eta)^{|I|-\ell} \tilde{\nu}_\Gamma(\{\gamma = (i_1, \ldots, i_k) \in \Gamma_i \mid i_\ell = i\} \times \Theta) > 0$.

Proof. For $\tau \in \{|I|, |I|+1, \ldots\}$ and for $S \subset I \setminus \{i\}$, we have

$$\begin{aligned}
&\pi(\{j \neq i \mid t_j < \tau\} = S, \theta | t_i = \tau) \\
&\quad = \pi(t_i = \tau, \{j \neq i \mid t_j < \tau\} = S, \theta)/\pi(t_i = \tau) \\
&\quad = (1-\varepsilon)\eta(1-\eta)^{\tau-|S|-1} \tilde{\nu}_\Gamma(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = \mathbf{1}_S\} \times \{\theta\})/\pi(t_i = \tau) \\
&\quad = (1-\eta)^{|I|-|S|-1} \tilde{\nu}_\Gamma(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = \mathbf{1}_S\} \times \{\theta\})/C_i,
\end{aligned}$$

as claimed. □

Claim A.3. *For any $i \in I$ such that $\tilde{\nu}_\Gamma(\Gamma_i \times \Theta) < 1$,*

$$\pi(\{j \neq i \mid t_j < \infty\} = S, \theta | t_i = \infty) = \nu(\mathbf{1}_S, \theta)/D_i$$

for all $S \subset I \setminus \{i\}$, where $D_i = (1-\varepsilon)(1 - \tilde{\nu}_\Gamma(\Gamma_i \times \Theta)) > 0$.

Proof. For $S \subset I \setminus \{i\}$, we have

$$\begin{aligned}
&\pi(\{j \neq i \mid t_j < \infty\} = S, \theta | t_i = \infty) \\
&\quad = \pi(t_i = \infty, \{j \neq i \mid t_j < \infty\} = S, \theta)/\pi(t_i = \infty) \\
&\quad = (1-\varepsilon)\tilde{\nu}_\Gamma(\{\gamma \in \Gamma \mid a(\gamma) = \mathbf{1}_S\} \times \{\theta\})/D_i \\
&\quad = \nu_\Gamma(\{\gamma \in \Gamma \mid a(\gamma) = \mathbf{1}_S\} \times \{\theta\})/D_i = \nu(\mathbf{1}_S, \theta)/D_i,
\end{aligned}$$

as claimed, where $(1-\varepsilon)\tilde{\nu}_\Gamma(\gamma, \theta) = \nu_\Gamma(\gamma, \theta)$ whenever $a(\gamma) = \mathbf{1}_S$. □

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

$$\begin{aligned}
&\sum_{S \subset I \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \tau\} = S, \theta | t_i = \tau) d_i(\mathbf{1}_S, \theta) \\
&= \sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I|-n(a_{-i}(\gamma))-1} \tilde{\nu}_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta)/C_i > 0,
\end{aligned}$$

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 $\underline{\sigma}_i$ be the pure strategy such that $\underline{\sigma}_i(t_i) = 1$ if and only if $t_i < \infty$. For a player i (with $\tilde{\nu}_\Gamma(\Gamma_i \times \Theta) < 1$) of type $t_i = \infty$, against $\underline{\sigma}_{-i}$ the expected payoff is

$$\sum_{S \subset I \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \infty\} = S, \theta | t_i = \infty) d_i(\mathbf{1}_S, \theta)$$
$$= \sum_{a_{-i} \in A_{-i}, \theta \in \Theta} \nu((0, a_{-i}), \theta) d_i(a_{-i}, \theta)/D_i \leq 0,$$

where the equality is by Claim A.3 and the inequality by (lower) obedience, which implies that playing 0 is a best response to $\underline{\sigma}_{-i}$. It therefore follows that $\underline{\sigma}$ is indeed the smallest equilibrium. Finally, by construction, $\underline{\sigma}$ induces ν, as desired.

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

$$D(\nu_\Gamma, \lambda) = \sum_{i \in I} \lambda_i \sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta)$$
$$= \sum_{\gamma \in \Gamma, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta)$$
$$= \sum_{a \in A, \theta \in \Theta} \sum_{\gamma : a(\gamma) = a} \nu_\Gamma(\gamma, \theta) \sum_{i \in S(a)} \lambda_i d_i(a_{-i}(\gamma), \theta),$$

which is linear in each of ν_Γ and λ, where for $\gamma \in \Gamma$, $S(\gamma)$ denotes the set of players that appear in γ.

First, ν satisfies sequential obedience (resp. weak sequential obedience) if and only if there exists $\nu_\Gamma \in N_\Gamma(\nu)$ such that $D(\nu_\Gamma, \lambda) >$ (resp. $\geq$) 0 for all $\lambda \in \Lambda(\nu)$, which in turn is equivalent to

$$\max_{\nu_\Gamma \in N_\Gamma(\nu)} \min_{\lambda \in \Lambda(\nu)} D(\nu_\Gamma, \lambda) > \text{(resp. } \geq \text{) } 0. \tag{A.4}$$

Second, (LHS of (3.8)) $= \max_{\nu_\Gamma \in N_\Gamma(\nu)} D(\nu_\Gamma, \lambda)$ for each $\lambda \in \Lambda(\nu)$. Hence, ν satisfies condition (3.8) if and only if

$$\min_{\lambda \in \Lambda(\nu)} \max_{\nu_\Gamma \in N_\Gamma(\nu)} D(\nu_\Gamma, \lambda) > \text{(resp. } \geq \text{) } 0. \tag{A.5}$$

Now, by the minimax theorem, we have $\max_{\nu_\Gamma} \min_\lambda D(\nu_\Gamma, \lambda) = \min_\lambda \max_{\nu_\Gamma} D(\nu_\Gamma, \lambda)$, 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 Φ. By Proposition 2, it suffices to show that $\nu \in \Delta(A \times \Theta)$ satisfies condition (3.8) in Proposition 2 if and only if it satisfies condition (3.12) in Proposition 3.

The "only if" part: Suppose that ν satisfies sequential obedience (resp. weak sequential obedience) and hence condition (3.8). Fix any $a \in A$ such that $S(a) \subsetneq I(\nu)$. Define $(\lambda_i^a)_{i\in I} \in \mathbb{R}_+^I$ by $\lambda_i^a = 1$ if $i \in I \setminus S(a)$ and $\lambda_i^a = 0$ if $i \in S(a)$. Note that $\lambda_i^a > 0$ for some $i \in I(\nu)$.

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

$$\begin{aligned}\max_{\gamma:a(\gamma)=a'} \sum_{i\in S(a')} \lambda_i^a d_i(a_{-i}(\gamma), \theta) &= \sum_{\ell=|S(a')\cap S(a)|+1}^{|S(a')|} (\Phi((1, a_{-i_\ell}(\gamma')), \theta) - \Phi((0, a_{-i_\ell}(\gamma')), \theta)) \\ &= \Phi(a', \theta) - \Phi(a \wedge a', \theta).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\Phi_\nu(\mathbf{1}) - \Phi_\nu(a) &= \sum_{a'\in A, \theta\in\Theta} \nu(a', \theta)(\Phi(a', \theta) - \Phi(a \wedge a', \theta)) \\ &= \sum_{a'\in A, \theta\in\Theta} \nu(a', \theta) \max_{\gamma:a(\gamma)=a'} \sum_{i\in S(a')} \lambda_i^a d_i(a_{-i}(\gamma), \theta),\end{aligned}$$

which is positive (resp. nonnegative) by condition (3.8).

The "if" part: Suppose that ν satisfies condition (3.12). We want to show that ν satisfies condition (3.8). Fix any $(\lambda_i)_{i\in I} \in \mathbb{R}_+^I$ such that $\lambda_i > 0$ for some $i \in I(\nu)$. Let $\gamma^\lambda = (i_1, \dots, i_{|I|})$ be a permutation of all players such that $\{i_1, \dots, i_{|I(\nu)|}\} = I(\nu)$ and $\lambda_{i_1} \le \cdots \le \lambda_{i_{|I(\nu)|}}$. Then we have

$$\begin{aligned}&(\text{LHS of (3.8)}) \\ &\ge \sum_{a'\in A, \theta\in\Theta} \nu(a', \theta) \sum_{i\in S(a')} \lambda_i(\Phi((1, a_{-i}(\gamma^\lambda)), \theta) - \Phi((0, a_{-i}(\gamma^\lambda)), \theta)) \\ &\ge \sum_{i\in I} \lambda_i \sum_{a'\in A, \theta\in\Theta} \nu(a', \theta)(\Phi((1, a_{-i}(\gamma^\lambda)) \wedge a', \theta) - \Phi((0, a_{-i}(\gamma^\lambda)) \wedge a'), \theta)) \\ &= \sum_{i\in I} \lambda_i(\Phi_\nu(1, a_{-i}(\gamma^\lambda)) - \Phi_\nu(0, a_{-i}(\gamma^\lambda))) \\ &= \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}}) \sum_{\ell=k}^{|I|} (\Phi_\nu(1, a_{-i_\ell}(\gamma^\lambda)) - \Phi_\nu(0, a_{-i_\ell}(\gamma^\lambda))) \\ &= \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}})(\Phi_\nu(\mathbf{1}) - \Phi_\nu(\mathbf{1}_{\{i_1,\dots,i_{k-1}\}})),\end{aligned}$$

which is positive (resp. nonnegative) by condition (3.12) 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 condition (3.12) in a potential game. The "only if" part is obvious. The "if" direction follows from convexity of Φ since for a perfect coordination outcome ν, we have

$$\begin{aligned}\Phi_\nu(\mathbf{1}) - \Phi_\nu(a) &= \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)(\Phi(\mathbf{1},\theta) - \Phi(a,\theta)) \\ &\geq \left(1 - \frac{n(a)}{|I|}\right) \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)\Phi(\mathbf{1},\theta) = \left(1 - \frac{n(a)}{|I|}\right) \Phi_\nu(\mathbf{1}) > (\text{resp. } \geq)\ 0\end{aligned}$$

for any $a \neq \mathbf{1}$.

A.5. **Proof of Theorem 2.** Suppose that Φ satisfies convexity and V satisfies restricted convexity with respect to Φ. As already noted, ν^* satisfies consistency (4.1b), weak sequential obedience (4.1c), and obedience, and hence is in $\overline{SI}$.

First, we show that $(\nu^*(\mathbf{1},\theta))_{\theta\in\Theta}$ is an optimal solution to the problem (4.1). Let $(\nu(\mathbf{1},\theta))_{\theta\in\Theta}$ be such that $0 \leq \nu(\mathbf{1},\theta) \leq \mu(\theta)$ and $\sum_{\theta\in\Theta} \nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta) < \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)V(\mathbf{1},\theta)$. Define $\xi = (\xi(\theta))_{\theta\in\Theta}$, $\xi^* = (\xi^*(\theta))_{\theta\in\Theta}$, and $\xi^{**} = (\xi^{**}(\theta))_{\theta\in\Theta}$ by $\xi(\theta) = \nu(\mathbf{1},\theta)V(\mathbf{1},\theta)$ for all $\theta \in \Theta$, $\xi^*(\theta) = \nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta)$ for all $\theta \in \Theta$, and $\xi^{**}(\theta^*) = \xi^*(\theta^*) + \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)V(\mathbf{1},\theta) - \sum_{\theta\in\Theta} \nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta) > \xi^*(\theta^*)$ and $\xi^{**}(\theta) = \xi^*(\theta)$ for all $\theta \neq \theta^*$.

Since $\sum_{\theta'\geq\theta} \xi(\theta') \leq \sum_{\theta'\geq\theta} \xi^{**}(\theta')$ for all $\theta \in \Theta$ and $\sum_{\theta\in\Theta} \xi(\theta) = \sum_{\theta\in\Theta} \xi^{**}(\theta)$ by the construction of ν^* and $\frac{\Phi(\mathbf{1},\theta)}{V(\mathbf{1},\theta)}$ is nondecreasing in θ, we have

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

But we have

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

since the first term in the right hand side of the equality equals 0 by (4.3), and $\Phi(\mathbf{1},\theta^*) < 0$. This means that $(\nu(\mathbf{1},\theta))_{\theta\in\Theta}$ is not feasible. This implies that $(\nu^*(\mathbf{1},\theta))_{\theta\in\Theta}$ is an optimal solution to the problem (4.1).

Next, we show that ν^* is an optimal outcome of the adversarial information design problem. For this, it suffices to show that for any outcome $\nu \in \overline{SI}$, there exists a perfectly coordinated outcome ν' that satisfies the constraints of consistency (4.1b) and weak sequential obedience (4.1c) and whose value is no smaller than that of ν. For each

(a,θ), define $\alpha(a,\theta) \in [0,1]$ by

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

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

Take any $\nu \in \overline{SI}$. By Corollary 1 and Proposition 3, ν satisfies consistency and condition (3.12) in Proposition 3. Define $\nu' \in \Delta(A \times \Theta)$ by

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

which satisfies the perfect coordination property. Since ν is consistent with μ, so is ν'. Since ν satisfies condition (3.12), we also have

$$\begin{aligned} \sum_{\theta \in \Theta} \nu'(\mathbf{1},\theta)\Phi(\mathbf{1},\theta) &= \sum_{a \in A, \theta \in \Theta} \alpha(a,\theta)\nu(a,\theta)\Phi(\mathbf{1},\theta) \\ &\geq \sum_{a \in A, \theta \in \Theta} \nu(a,\theta)\Phi(a,\theta) = \Phi_\nu(\mathbf{1}) \geq 0. \end{aligned}$$

Therefore, ν' satisfies (4.1c). For the value of the objective function, we have

$$\begin{aligned} \sum_{\theta \in \Theta} \nu'(\mathbf{1},\theta)V(\mathbf{1},\theta) &= \sum_{a \in A, \theta \in \Theta} \nu(a,\theta)\alpha(a,\theta)V(\mathbf{1},\theta) \\ &\geq \sum_{a \in A, \theta \in \Theta} \nu(a,\theta)V(a,\theta). \end{aligned}$$

This completes the proof of Theorem 2.

References

Arieli, I. and Y. Babichenko (2019): "Private Bayesian Persuasion," *Journal of Economic Theory*, 182, 185–217.

Aumann, R. J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96.

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games with Incomplete Information," *Econometrica*, 81, 1251–1308.

——— (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures," *Theoretical Economics*, 11, 487–522.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Carlsson, H. and E. van Damme (1993): "Global Games and Equilibrium Selection," *Econometrica*, 61, 989–1018.

Carroll, G. (2016): "Informationally Robust Trade and Limits to Contagion," *Journal of Economic Theory*, 166, 334–361.

Cooper, R. (1994): "Equilibrium Selection in Imperfectly Competitive Economies with Multiple Equilibria," *Economic Journal*, 104, 1106–1122.

Doval, L. and J. Ely (2020): "Sequential Information Design," *Econometrica*, 88, 2575–2608.

Edmond, C. (2013): "Information Manipulation, Coordination and Regime Change," *Review of Economic Studies*, 80, 1422–58.

Forges, F. (1993): "Five Legitimate Definitions of Correlated Equilibrium in Games with Incomplete Information," *Theory and Decision*, 35, 277–310.

Frankel, D. M., S. Morris, and A. Pauzner (2003): "Equilibrium Selection in Global Games with Strategic Complementarities," *Journal of Economic Theory*, 108, 1–44.

Galperti, S. and J. Perego (2020): "Information Systems," Mimeo.

Goldstein, I. and A. Pauzner (2005): "Demand-Deposit Contracts and the Probability of Bank Runs," *Journal of Finance*, 60, 1293–1327.

Gossner, O. and R. Veiel (2022): "Rationalizable Outcomes in Games with Incomplete Information," Mimeo.

Halac, M., E. Lipnowski, and D. Rappoport (2021): "Rank Uncertainty in Organizations," *American Economic Review*, 111, 757–86.

——— (2022): "Addressing Strategic Uncertainty with Incentives and Information," *AEA Papers and Proceedings*, 112, 431–437.

Harsanyi, J. C. and R. Selten (1988): *A General Theory of Equilibrium Selection in Games*, Cambridge: MIT Press.

Hoshino, T. (2022): "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," *International Economic Review*, 63, 755–776.

Inostroza, N. and A. Pavan (2022): "Adversarial Coordination and Public Information Design," Mimeo.

Kajii, A. and S. Morris (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65, 1283–1309.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kamien, M. I., Y. Tauman, and S. Zamir (1990): "On the Value of Information in a Strategic Conflict," *Games and Economic Behavior*, 2, 129–153.

Leister, C. M., Y. Zenou, and J. Zhou (2022): "Social Connectedness and Local Contagion," *Review of Economic Studies*, 89, 372–410.

Li, F., Y. Song, and M. Zhao (2022): "Global Manipulation by Local Obfuscation," Mimeo.

Mathevet, L., J. Perego, and I. Taneva (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

Mathevet, L. and I. Taneva (2022): "Organized Information Transmission," Mimeo.

Mertens, J.-F. and S. Zamir (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

Milgrom, P. and J. Roberts (1990): "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255–1277.

Monderer, D. and D. Samet (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1, 170–190.

Monderer, D. and L. S. Shapley (1996): "Potential Games," *Games and Economic Behavior*, 14, 124–143.

Moriya, F. and T. Yamashita (2020): "Asymmetric-Information Allocation to Avoid Coordination Failure," *Journal of Economics & Management Strategy*, 29, 173–186.

Morris, S., D. Oyama, and S. Takahashi (2022a): "Implementation via Information Design using Global Games," SSRN 4140792.

——— (2022b): "On the Joint Design of Information and Transfers," SSRN 4156831.

Morris, S. and H. S. Shin (1998): "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 88, 587–597.

——— (2004): "Coordination Risk and the Price of Debt," *European Economic Review*, 48, 133–153.

——— (2007): "Common Belief Foundations of Global Games," Mimeo.

Morris, S., H. S. Shin, and M. Yildiz (2016): "Common Belief Foundations of Global Games," *Journal of Economic Theory*, 163, 826–848.

Morris, S. and T. Ui (2005): "Generalized Potentials and Robust Sets of Equilibria," *Journal of Economic Theory*, 124, 45–78.

——— (2020): "Incomplete Information Robustness," Mimeo.

Myerson, R. (1991): *Game Theory: Analysis of Conflict*, Harvard University Press.

Oyama, D. and S. Takahashi (2020): "Generalized Belief Operator and Robustness in Binary-Action Supermodular Games," *Econometrica*, 88, 693–726.

Rubinstein, A. (1989): "The Electronic Mail Game: Strategic Behavior under 'Almost Common Knowledge'," *American Economic Review*, 79, 385–391.

Sandmann, C. (2020): "Recursive Information Design," Mimeo.

Segal, I. (2003): "Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?" *Journal of Economic Theory*, 113, 147–181.

Taneva, I. (2019): "Information Design," *American Economic Journal: Microeconomics*, 11, 151–85.

Ui, T. (2001): "Robust Equilibria of Potential Games," *Econometrica*, 69, 1373–1380.

Vives, X. (1990): "Nash Equilibrium with Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305–321.

Winter, E. (2004): "Incentives and Discrimination," *American Economic Review*, 94, 764–773.

B.1. **Details of the Leading Example.** In this section, we analyze in more detail the example from Section 2 in the main text. Recall that payoffs were given by the following tables, where player 1 is the row player and player 2 is the column player:

b	Not	Invest
Not	$0, 0$	$0, -8$
Invest	$-7, 0$	$-4, -5$

g	Not	Invest
Not	$0, 0$	$0, 1$
Invest	$2, 0$	$5, 4$

.

The game is a special case of the investment game which has a potential (Example 1), and the potential of this game satisfies convexity (Example 3). The designer wants to maximize the expected number of players who choose action 1, i.e., $V(a, \theta) = n(a)$ for all $a \in A$ and $\theta \in \Theta$, so that restricted convexity is satisfied.

As a benchmark, we first study the case where the designer can only send public signals. When the information structure is generated by public signals, the players share the same posterior belief over $\Theta = \{\mathbf{b}, \mathbf{g}\}$. Let q denote the posterior probability of $\mathbf{g}$. Given q, the average game is given by

	Not	Invest
Not	$0, 0$	$0, 9q - 8$
Invest	$9q - 7, 0$	$9q - 4, 9q - 5$

, (B.1)

which has a convex potential

	Not	Invest
Not	$0, 0$	$9q - 8$
Invest	$9q - 7$	$18q - 12$

.

Note that due to the convexity of the potential, pure equilibria of this game are always either (Not Invest, Not Invest) or (Invest, Invest). The profile (Invest, Invest) is an equilibrium if and only if $q \geq \frac{5}{9}$, while it is the smallest (hence unique) equilibrium if and only if $q > \frac{7}{9}$. By a concavification argument from Bayesian persuasion, the optimal value under partial implementation with public signals as a function of $\mu(\mathbf{g})$ is given by the solid line segments in Figure 1 with $q^* = \frac{5}{9}$ and $V^* = \frac{9}{5}$. Similarly, that under S-implementation with public signals is given in Figure 1 by substituting $q^* = \frac{7}{9}$ and $V^* = \frac{9}{7}$. When $\mu(\mathbf{g}) = \frac{1}{2}$ $(= \frac{1}{10} \times 0 + \frac{9}{10} \times \frac{5}{9} = \frac{5}{14} \times 0 + \frac{9}{14} \times \frac{7}{9})$ as in Section 2, the optimal

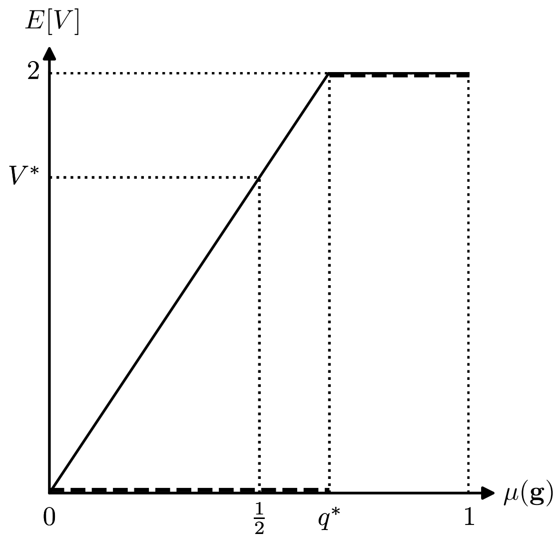

FIGURE 1. Optimal values—concavification

partially implementable outcome with public signals is

$\mathbf{b}$	Not	Invest
Not	$\frac{1}{10}$	0
Invest	0	$\frac{2}{5}$

$\mathbf{g}$	Not	Invest
Not	0	0
Invest	0	$\frac{1}{2}$

which gives the value $\frac{9}{5} = 1.8$. Note that this is a perfect coordination outcome, where only the obedience condition of player 2 (less willing to invest) binds and the slack in that of player 1 (more willing to invest) can be exploited to induce more investment by player 1 if private signals can be used. The optimal outcome under S-implementation with public signals is approached, as $\delta \to 0$, by

$\mathbf{b}$	Not	Invest
Not	$\frac{5}{14} + \delta$	0
Invest	0	$\frac{1}{7} - \delta$

$\mathbf{g}$	Not	Invest
Not	0	0
Invest	0	$\frac{1}{2}$

with the value arbitrarily close to $\frac{9}{7} \approx 1.3$, which is S-implemented (in fact fully implemented) by the direct information structure. Indeed, it is induced, for example, by the ordered outcome ν_Γ such that $\nu_\Gamma(\emptyset, \mathbf{b}) = \frac{5}{14} + \delta$, $\nu_\Gamma(12, \mathbf{b}) = \frac{1}{7} - \delta$, and $\nu_\Gamma(12, \mathbf{g}) = \frac{1}{2}$ (and $\nu_\Gamma(\gamma, \theta) = 0$ otherwise) which satisfies the "public sequential obedience" condition $\sum_{\theta \in \Theta} \nu_\Gamma(12, \theta) d_i(a_{-i}(12), \theta) > 0$ for all $i \in I$, where in the limit as $\delta \to 0$, only the condition for player 1 binds.

Now, (Invest, Invest) is a (weakly) risk dominant equilibrium, or equivalently a (weak) potential maximizer, in the average game (B.1) if and only if $q \geq \frac{2}{3}$. Indeed, if $\mu(\mathbf{g}) =$

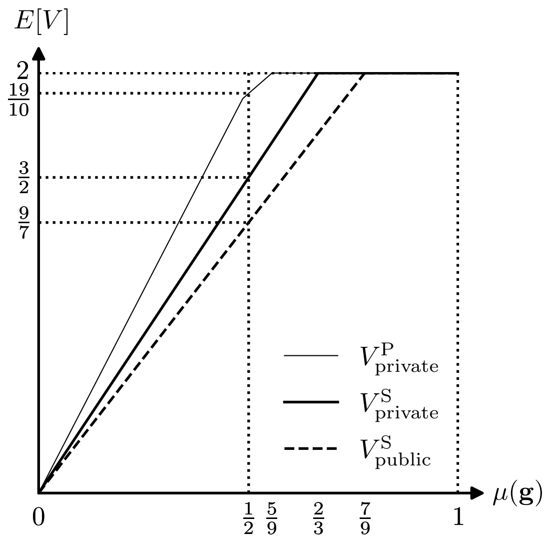

FIGURE 2. Optimal values—comparison

$q \geq \frac{2}{3}$, the ordered outcome given by

	$\mathbf{b}$	$\mathbf{g}$
12	$\frac{2}{3}(1-q)$	$\frac{2}{3}q$
21	$\frac{1}{3}(1-q)$	$\frac{1}{3}q$

(and $\nu_\Gamma(\gamma, \theta) = 0$ otherwise) satisfies weak sequential obedience. When $\mu(\mathbf{g}) = \frac{1}{2}$ $(= \frac{1}{4} \times 0 + \frac{3}{4} \times \frac{2}{3})$, the $\frac{1}{4}$-$\frac{3}{4}$ convex combination of the ordered outcome that assigns probability 1 to $(\gamma, \theta) = (\emptyset, \mathbf{b})$ and the above ordered outcome with $q = \frac{2}{3}$ satisfies weak sequential obedience and induces the optimal outcome (2.3) under S-implementation with private signals, as shown in Section 3.1, with the value arbitrarily close to $\frac{3}{2} = 1.5$ (let $q^* = \frac{2}{3}$ and $V^* = \frac{3}{2}$ in Figure 1).

Finally, the optimal outcome under partial implementation with private signals for $\mu(\mathbf{g}) = q < \frac{5}{9}$ is given by the following. If $q < \frac{35}{72}$, then

$\mathbf{b}$	Not	Invest
Not	$1 - \frac{72}{35}q$	0
Invest	$\frac{9}{35}q$	$\frac{4}{5}q$

$\mathbf{g}$	Not	Invest
Not	0	0
Invest	0	q

with value $\frac{27}{7}q$; and if $\frac{35}{72} \leq q < \frac{5}{9}$, then

$\mathbf{b}$	Not	Invest
Not	0	0
Invest	$1 - \frac{9}{5}q$	$\frac{4}{5}q$

$\mathbf{g}$	Not	Invest
Not	0	0
Invest	0	q

with value $1 + \frac{9}{5}q$. Figure 2 depicts the optimal values under partial implementation with private signals $V^{\mathrm{P}}_{\mathrm{private}}$ (thin solid line), S-implementation with private signals $V^{\mathrm{S}}_{\mathrm{private}}$ (thick solid line), and S-implementation with public signals $V^{\mathrm{S}}_{\mathrm{public}}$ (dashed line). (The omitted graph of the optimal value under partial implementation with public signals is obtained by the line that connects the origin and the point $(\frac{5}{9}, 2)$.)

B.2. **Omitted Proofs.**

B.2.1. *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 $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$. Let $\underline{\nu} \in \Delta(A \times \Theta)$ be any outcome that satisfies consistency, obedience, sequential obedience with, say, $\underline{\nu}_\Gamma \in \Delta(\Gamma \times \Theta)$, and grain of dominance.[31] Then define $\nu^\varepsilon \in \Delta(A \times \Theta)$ by $\nu^\varepsilon = (1-\varepsilon)\nu + \varepsilon\underline{\nu}$. Clearly, ν^ε satisfies consistency, obedience, sequential obedience with $(1-\varepsilon)\nu_\Gamma + \varepsilon\underline{\nu}_\Gamma$, 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 $\nu \in \overline{SI}$.

B.2.2. *Proof of Corollary 2.* First, we claim that for any $\nu \in \Delta(A \times \Theta)$ that satisfies consistency, sequential obedience, and grain of dominance, there exists $\hat{\nu} \in SI$ that first-order stochastically dominates ν. Indeed, given such an outcome ν, 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 ν.

Now 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 ν that satisfy consistency, sequential obedience, and grain of dominance: for example, let $\nu^\varepsilon = (1-\varepsilon)\nu + \varepsilon\underline{\nu}$ with an outcome $\underline{\nu}$ as in the proof of Corollary 1. Then, as claimed above, for each ε, there exists an outcome $\hat{\nu}^\varepsilon \in SI$ that first-order stochastically dominates ν^ε. Then a limit point of $\hat{\nu}^\varepsilon$, which is contained in $\overline{SI}$, first-order stochastically dominates ν.

B.2.3. *Proof of Corollary 4.* Part (1): The "only if" part follows from Theorem 1(1) by a continuity argument. To prove the "if" part, let $\xi \in \Delta(A)$ satisfy obedience and weak

[31]For example, let $\underline{\nu}$ 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 assumptions of dominance state and full support.

sequential obedience with an ordered outcome $\rho \in \Delta(\Gamma)$ in $(d_i(\cdot, \theta^*))_{i\in I}$. Let $\bar{\rho} \in \Delta(\Gamma)$ be an ordered outcome that satisfies sequential obedience in $(d_i(\cdot, \overline{\theta}))_{i\in I}$, where $\bar{\rho}(\{\gamma \in \Gamma \mid a(\gamma) = \mathbf{1}\}) = 1$ by the dominance state assumption. Then for each k, define $\mu^k \in \Delta(\Theta)$ by $\mu^k(\theta^*) = 1 - \frac{1}{k}$ and $\mu^k(\overline{\theta}) = \frac{1}{k}$, and $\nu_\Gamma^k \in \Delta(\Gamma \times \Theta)$ by $\nu_\Gamma^k(\cdot, \theta^*) = \left(1 - \frac{1}{k}\right) \rho$ and $\nu_\Gamma^k(\cdot, \overline{\theta}) = \frac{1}{k}\bar{\rho}$, and let $\nu^k \in \Delta(A \times \Theta)$ be the outcome induced by ν_Γ^k. Clearly, ν^k is consistent with μ^k and satisfies obedience, sequential obedience, and grain of dominance. Hence, $\nu^k \in SI(\mu^k)$ by Theorem 1(2). Since $\mu^k(\theta^*) \to 1$ and $\sum_{\theta\in\Theta} \nu^k(\cdot, \theta) \to \xi$ as $k \to \infty$, ξ is limit S-implementable at θ^*.

Part (2): Let $\xi \in \Delta(A)$ satisfy weak sequential obedience. Then, as in the proof of part (1), there exist a sequence of priors $\mu^k \in \Delta(\Theta)$ such that $\mu^k(\theta^*) \to 1$ as $k \to \infty$ and a sequence of outcomes $\nu^k \in \Delta(A \times \Theta)$ converging to ξ each of which is consistent with μ^k and satisfies sequential obedience and grain of dominance. By Corollary 2, for each k, there exists $\hat{\nu}^k \in SI(\mu^k)$ that first-order stochastically dominates ν^k. Then a limit point of $\sum_{\theta\in\Theta} \hat{\nu}^k(\cdot, \theta)$ is limit S-implementable at θ^* and first-order stochastically dominates ξ.

B.3. **Full Implementation.** In this section, we study 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 optimal information design problem under S-implementation is equivalent to that under full implementation.

To proceed, we add a symmetric dominance state assumption that there exists $\underline{\theta} \in \Theta$ such that $d_i(\mathbf{1}_{-i}, \underline{\theta}) < 0$ 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 $a^0(\gamma^0) = \mathbf{1} - a(\gamma^0) \in A$ for the action profile such that player i plays action 0 if and only if i is listed in γ^0 and $a^0_{-i}(\gamma^0) \in A_{-i}$ for the action profile such that only players before i in γ^0 play action 0. Thus an ordered outcome $\nu_\Gamma^0 \in \Delta(\Gamma \times \Theta)$ *reverse induces* $\nu \in \Delta(A \times \Theta)$ if

$$\nu(a, \theta) = \sum_{\gamma^0 :\, a^0(\gamma^0) = a} \nu_\Gamma^0(\gamma^0, \theta).$$

Definition B.1. An ordered outcome $\nu_\Gamma^0 \in \Delta(\Gamma \times \Theta)$ satisfies *reverse sequential obedience* if

$$\sum_{\gamma^0 \in \Gamma_i, \theta\in\Theta} \nu_\Gamma^0(\gamma^0, \theta) d_i(a^0_{-i}(\gamma^0), \theta) < 0 \qquad \text{(B.2)}$$

for all $i \in I$ such that $\nu_\Gamma^0(\Gamma_i \times \Theta) > 0$. An outcome $\nu \in \Delta(A \times \Theta)$ satisfies *reverse sequential obedience* if there exists an ordered outcome $\nu_\Gamma^0 \in \Delta(\Gamma \times \Theta)$ that reverse induces ν and satisfies reverse sequential obedience.

Definition B.2. Outcome ν satisfies *two-sided grain of dominance* if $\nu(\mathbf{1}, \overline{\theta}) > 0$ and $\nu(\mathbf{0}, \underline{\theta}) > 0$.

Theorem B.1. (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 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 Section B.3.1 below, 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.[32] Yet, we show that *maximal* S-implementable outcomes (with respect to first-order stochastic dominance) must indeed be fully implementable.

Proposition B.1. *For any $\nu \in \overline{SI}$, there exists $\hat{\nu} \in \overline{FI}$ that first-order stochastically dominates ν.*

The proof is given in Section B.3.3. This proposition, in particular, implies that $FI \neq \emptyset$.

[32]For 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.

Consider, as in Section 1.4, the optimal information design problem but with full implementation:

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

By Proposition B.1, under the monotonicity assumption on V, solving this problem amounts to solving the problem with S-implementation: we have

$$\max_{\nu\in\overline{FI}}\sum_{a\in A,\theta\in\Theta}\nu(a,\theta)V(a,\theta)=\max_{\nu\in\overline{SI}}\sum_{a\in A,\theta\in\Theta}\nu(a,\theta)V(a,\theta),$$

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

B.3.1. *Proof of Theorem B.1(2).* In the following, for $S\subset I$, we denote by $\Gamma(S)\subset\Gamma$ the set of sequences of distinct players in S and by $\Pi(S)\subset\Gamma(S)$ the set of permutations of all players in S.

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

$$\tilde{\nu}_\Gamma^+(\gamma,\theta)=\begin{cases}\dfrac{\nu_\Gamma^+(\gamma,\theta)-\varepsilon}{1-2\varepsilon} & \text{if } (\gamma,\theta)=(\overline{\gamma},\overline{\theta}),(\emptyset,\underline{\theta}),\\ \dfrac{\nu_\Gamma^+(\gamma,\theta)}{1-2\varepsilon} & \text{otherwise,}\end{cases}$$

and

$$\tilde{\nu}_\Gamma^-(\gamma,\theta)=\begin{cases}\dfrac{\nu_\Gamma^-(\gamma,\theta)-\varepsilon}{1-2\varepsilon} & \text{if } (\gamma,\theta)=(\underline{\gamma},\underline{\theta}),(\emptyset,\overline{\theta}),\\ \dfrac{\nu_\Gamma^-(\gamma,\theta)}{1-2\varepsilon} & \text{otherwise,}\end{cases}$$

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

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

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

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

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

$$\tilde{\nu}(a,\theta) = \begin{cases} \dfrac{\nu(a,\theta)-\varepsilon}{1-2\varepsilon} & \text{if } (a,\theta) = (\mathbf{1},\overline{\theta}), (\mathbf{0},\underline{\theta}), \\ \dfrac{\nu(a,\theta)}{1-2\varepsilon} & \text{otherwise.} \end{cases}$$

Observe that $\sum_{\gamma^+:a(\gamma^+)=a} \tilde{\nu}_\Gamma^+(\gamma^+,\theta) = \sum_{\gamma^-:a^0(\gamma^-)=a} \tilde{\nu}_\Gamma^-(\gamma^-,\theta) = \tilde{\nu}(a,\theta)$ for all $(a,\theta) \in A \times \Theta$.

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

$$\bar{q} d_i(\mathbf{0}_{-i},\overline{\theta}) + (1-\bar{q}) \min_{\theta \neq \overline{\theta}} d_i(\mathbf{0}_{-i},\theta) > 0,$$
$$\bar{q} d_i(\mathbf{1}_{-i},\underline{\theta}) + (1-\bar{q}) \max_{\theta \neq \underline{\theta}} d_i(\mathbf{1}_{-i},\theta) < 0$$

for all $i \in I$. Then let $\eta > 0$ be such that

$$\frac{\frac{\varepsilon}{|I|-1}}{\frac{\varepsilon}{|I|-1} + \eta} \geq \bar{q},$$

and

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I|-n(a_{-i}(\gamma))-1} \tilde{\nu}_\Gamma^+(\gamma,\theta) d_i(a_{-i}(\gamma),\theta) > 0$$

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

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I|-n^0(a^0_{-i}(\gamma))-1} \tilde{\nu}_\Gamma^-(\gamma,\theta) d_i(a^0_{-i}(\gamma),\theta) < 0$$

for all $i \in I$ such that $\tilde{\nu}_\Gamma^-(\Gamma_i \times \Theta) > 0$, where $n^0(a^0_{-i}(\gamma))$ is the number of players playing action 0 in the action profile $a^0_{-i}(\gamma)$.

Now construct the type space (T,π) as follows. For each $i \in I$, let $T_i = \{1,2,\ldots\} \times A_i$. 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

$$\pi(t,\theta) = \begin{cases} (1-2\varepsilon)\eta(1-\eta)^m \dfrac{\tilde{\nu}_\Gamma^+(\gamma^+,\theta)\tilde{\nu}_\Gamma^-(\gamma^-,\theta)}{\tilde{\nu}(a,\theta)} & \text{if } \tilde{\nu}(a,\theta) > 0 \text{ and there exist } m \in \mathbb{N}, \\ & \gamma^+ \in \Pi(S(a)), \text{ and } \gamma^- \in \Pi(I \setminus S(a)) \\ & \text{such that } s_i = m + \ell(i,\gamma^+) \text{ for all} \\ & i \in S(a) \text{ and } s_i = m + \ell(i,\gamma^-) \text{ for} \\ & \text{all } i \in I \setminus S(a), \\ \dfrac{\varepsilon}{|I|-1} & \text{if } 1 \leq s_1 = \cdots = s_{|I|} \leq |I|-1 \text{ and} \\ & (a,\theta) = (\mathbf{1},\overline{\theta}), (\mathbf{0},\underline{\theta}), \\ 0 & \text{otherwise,} \end{cases}$$

where $\ell(i,\gamma) = \ell$ if $i = i_\ell$. Observe that π is consistent with μ: $\sum_t \pi(t,\theta) = \mu(\theta)$ 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 ν, as desired.

B.3.2. *(Reverse) Sequential Obedience in Complete Information Games.* In this section, we report an important property of (reverse) sequential obedience in complete information games, which will be used in the proof of Proposition B.1 in Section B.3.3. A complete information BAS game is given by a profile of payoff difference functions $f_i \colon A_{-i} \to \mathbb{R}$, $i \in I$. Let $\overline{X} \subset \Delta(A)$ (resp. $\overline{X}^0 \subset \Delta(A)$) denote the set of outcomes that satisfy weak sequential obedience (resp. weak reverse sequential obedience) in $(f_i)_{i\in I}$, which is endowed with the first-order stochastic dominance order.

Proposition B.2. *In any complete information BAS game, the following hold:*

(1) $\overline{X}$ *has a largest element, which is degenerate on some action profile and satisfies reverse sequential obedience.*

(2) $\overline{X}^0$ *has a smallest element, which is degenerate on some action profile and satisfies sequential obedience.*

In particular, from this proposition it follows that any complete information BAS game has an action profile (or an outcome degenerate on some action profile) that satisfies weak sequential obedience and reverse sequential obedience (by part (1)) and an action profile that satisfies sequential obedience and weak reverse sequential obedience (by part (2)).

Proof. By symmetry, we only prove (2). By Oyama and Takahashi (2020, Lemma 2(2)) along with its convexity, $\overline{X}^0$ has a smallest element, which is degenerate on some action profile, say $a^1 \in A$. Denote $S^1 = S(a^1) = \{i \in I \mid a^1_i = 1\}$. Let $\mathbf{0}^0 \in \prod_{j\in I\setminus S^1} A_j$ denote the action profile of players in $I \setminus S^1$ such that all players in $I \setminus S^1$ play action 0, and for $i \in S^1$ and $\gamma^0 \in \Gamma(S^1)$, let $b^0_{-i}(\gamma^0) \in \prod_{j\in S^1\setminus\{i\}} A_j$ denote the action profile of players in S^1 such that only the players in S^1 that appear before i in γ^0 play action 0.

Claim B.1. *There exists no ordered outcome $\rho^0 \in \Delta(\Gamma(S^1) \setminus \{\emptyset\})$ such that*

$$\sum_{\gamma^0\in\Gamma(S^1)\setminus\{\emptyset\}} \rho^0(\gamma^0) f_i(b^0_{-i}(\gamma^0), \mathbf{0}^0) \leq 0 \tag{B.3}$$

for all $i \in S^1$.

Proof. Assume that there exists $\rho^0 \in \Delta(\Gamma(S^1) \setminus \{\emptyset\})$ that satisfies (B.3) for all $i \in S^1$. Let $\bar{\rho}^0 \in \Delta(\Pi(I \setminus S^1))$ be an ordered outcome that establishes weak reverse sequential obedience of a^1. Define $\hat{\rho}^0 \in \Delta(\Gamma)$ by

$$\hat{\rho}^0(\gamma) = \begin{cases} \bar{\rho}^0(\gamma_0^0)\rho^0(\gamma_1^0) & \text{if } \gamma^0 = (\gamma_0^0, \gamma_1^0) \text{ for some } \gamma_0^0 \in \Pi(I \setminus S^1) \text{ and } \gamma_1^0 \in \Gamma(S^1) \setminus \{\emptyset\}, \\ 0 & \text{otherwise.} \end{cases}$$

Then this $\hat{\rho}^0$ satisfies weak reverse sequential obedience and reverse induces an outcome that is strictly stochastically dominated by (the degenerate outcome on) a^1. The existence of such an ordered outcome contradicts the condition that a^1 is the smallest element of $\overline{X}^0$. □

By Claim B.1, it follows from a duality theorem (or from Oyama and Takahashi (2020, Lemma 2(1)) applied to the "subgame" $(f_i^1)_{i \in S^1}$ defined by $f_i^1(b_{-i}) = f_i(b_{-i}, \mathbf{0}^0)$ for $b_{-i} \in \prod_{j \in S^1 \setminus \{i\}}$) that there exists $(\lambda_i^1)_{i \in S^1} \in \mathbb{R}_{++}^{S^1}$ such that

$$\sum_{i \in S(\gamma^0)} \lambda_i^1 f_i(b_{-i}^0(\gamma^0), \mathbf{0}^0) > 0$$

for all $\gamma^0 \in \Gamma(S^1) \setminus \{\emptyset\}$. This is equivalent to the condition that for any $\gamma = (i_1, \ldots, i_{|S^1|}) \in \Pi(S^1)$,

$$\sum_{\ell=k}^{|S^1|} \lambda_{i_\ell}^1 f_{i_\ell}(a_{-i_\ell}(\gamma)) > 0 \tag{B.4}$$

for all $k = 1, \ldots, |S^1|$. We want to show that the condition (3.8) in Proposition 2 holds for the degenerate outcome on a^1 (with $|\Theta| = 1$). Fix any $(\lambda_i)_{i \in I} \in \mathbb{R}_+^I$ such that $\lambda_i > 0$ for some $i \in S^1$. Let $\gamma^\lambda = (i_1, \ldots, i_{|I|})$ be a permutation of all players such that $\{i_1, \ldots, i_{|S^1|}\} = S^1$ and $\frac{\lambda_{i_1}}{\lambda_{i_1}^1} \leq \cdots \leq \frac{\lambda_{i_{|S^1|}}}{\lambda_{i_{|S^1|}}^1}$. Then we have

$$\begin{aligned} (\text{LHS of (3.8)}) &\geq \sum_{i \in S^1} \lambda_i f_i(a_{-i}(\gamma^\lambda)) \\ &= \sum_{k=1}^{|S^1|} \left(\frac{\lambda_{i_k}}{\lambda_{i_k}^1} - \frac{\lambda_{i_{k-1}}}{\lambda_{i_{k-1}}^1} \right) \sum_{\ell=k}^{|S^1|} \lambda_{i_\ell}^1 f_{i_\ell}(a_{-i_\ell}(\gamma^\lambda)) > 0 \end{aligned}$$

by (B.4), where we set $\frac{\lambda_{i_0}}{\lambda_{i_0}^1} = 0$. Therefore, it follows from Proposition 2 that a^1 satisfies sequential obedience. □

B.3.3. *Proof of Proposition B.1.* By Proposition B.2, we have the following:

Lemma B.1. *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 ν and satisfies consistency, sequential obedience, reverse sequential obedience, and two-sided grain of dominance.*

Proof. For each $S \subset I$ and $\theta \in \Theta$, apply Proposition B.2(1) to the complete information game $(d_i((\cdot, \mathbf{1}_{I\setminus S}), \theta))_{i\in S}$: 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 $(d_i((\cdot, \mathbf{1}_{I\setminus S}), \theta))_{i\in S}$ with $\rho_{S,\theta}$ and $\rho^0_{S,\theta}$, respectively, where $\rho_{S,\theta}(\Pi(S(a^*_{S,\theta}))) = 1$ and $\rho^0_{S,\theta}(\Pi(S \setminus S(a^*_{S,\theta}))) = 1$, and, by convention, $\rho_{\emptyset,\theta}(\emptyset) = \rho^0_{\emptyset,\theta}(\emptyset) = 1$. By construction, for any $S \subset I$ and $\theta \in \Theta$, we have

$$\sum_{\gamma\in\Gamma(S)\cap\Gamma_i} \rho_{S,\theta}(\gamma) d_i(a_{-i}((\gamma', \gamma)), \theta) \geq 0 \tag{B.5}$$

for all $\gamma' \in \Pi(I \setminus S)$ and all $i \in S(a^*_{S,\theta})$, and

$$\sum_{\gamma^0\in\Gamma(S)\cap\Gamma_i} \rho^0_{S,\theta}(\gamma^0) d_i(a^0_{-i}(\gamma^0), \theta) < 0 \tag{B.6}$$

for all $i \in S \setminus S(a^*_{S,\theta})$. Note, in particular, that $a^*_{I,\underline{\theta}} = \mathbf{0}$ and hence $\rho_{I,\underline{\theta}}(\emptyset) = 1$ 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 $\hat{\nu}_\Gamma, \hat{\nu}^0_\Gamma \in \Delta(\Gamma \times \Theta)$ by

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

and

$$\hat{\nu}^0_\Gamma(\gamma^0, \theta) = \sum_{a:S(a)\subset I\setminus S(\gamma^0)} \nu(a, \theta) \rho^0_{I\setminus S(a),\theta}(\gamma^0),$$

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

$$\begin{aligned}\hat{\nu}(a, \theta) &= \sum_{\gamma:a(\gamma)=a} \hat{\nu}_\Gamma(\gamma, \theta) \\ &= \sum_{a':S(a')\subset S(a),\, S(a)\setminus S(a')=S\left(a^*_{I\setminus S(a'),\theta}\right)} \nu(a', \theta).\end{aligned}$$

One can verify that $\hat{\nu}$ satisfies consistency and two-sided grain of dominance and first-order stochastically dominates ν, and that $\hat{\nu}^0_\Gamma$ 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

$$\sum_{\gamma\in\Gamma_i,\theta\in\Theta} \hat{\nu}_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta)$$

$$= \sum_{\theta \in \Theta} \sum_{\gamma' \in \Gamma_i} \nu_\Gamma(\gamma', \theta) d_i(a_{-i}(\gamma'), \theta)$$
$$+ \sum_{\theta \in \Theta} \sum_{S \subset I \setminus \{i\}} \sum_{\gamma' \in \Pi(S)} \nu_\Gamma(\gamma', \theta) \sum_{\gamma'' \in \Gamma(I \setminus S) \cap \Gamma_i} \rho_{I \setminus S, \theta}(\gamma'') d_i(a_{-i}((\gamma', \gamma'')), \theta) > 0,$$

where the inequality follows from the sequential obedience of ν_Γ and (B.5).

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

$$\sum_{\gamma^0 \in \Gamma_i, \theta \in \Theta} \hat{\nu}_\Gamma^0(\gamma^0, \theta) d_i(a_{-i}^0(\gamma^0), \theta)$$
$$= \sum_{\theta \in \Theta} \sum_{S \subset I \setminus \{i\}} \sum_{\gamma' \in \Pi(S)} \nu_\Gamma(\gamma', \theta) \sum_{\gamma^0 \in \Gamma(I \setminus S) \cap \Gamma_i} \rho_{I \setminus S, \theta}^0(\gamma^0) d_i(a_{-i}^0(\gamma^0), \theta) < 0,$$

where the inequality follows from (B.6). □

From Lemma B.1, we have the following:

Lemma B.2. *If an outcome ν satisfies consistency and weak sequential obedience, then there exists an outcome $\hat{\nu} \in \overline{FI}$ that first-order stochastically dominates ν.*

Proof. First, it follows from Lemma B.1 and Theorem B.1(2) that for any $\nu \in \Delta(A \times \Theta)$ that satisfies consistency, sequential obedience, and two-sided grain of dominance, there exists $\hat{\nu} \in FI$ that first-order stochastically dominates ν.

Now 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 ν that satisfy consistency, sequential obedience, and two-sided grain of dominance. As noted above, for each ε, there exists an outcome $\hat{\nu}^\varepsilon \in FI$ that first-order stochastically dominates ν^ε. Then a limit point of $\hat{\nu}^\varepsilon$, which is contained in $\overline{FI}$, first-order stochastically dominates ν. □

Finally, Proposition B.1 follows from Corollary 1 and Lemma B.2.

B.4. **Alternative Assumptions.**

B.4.1. *Non-Supermodular Payoffs.* In this section, we consider general binary-action games with possibly non-supermodular payoffs and demonstrate that our arguments will still work if we employ rationalizability, rather than equilibrium, as a solution concept in implementing incomplete information games and strengthen the sequential obedience condition accordingly.

Let a base game $(d_i)_{i\in I}$ be given, which may not be supermodular. To simplify the argument, we focus on the "always play **1**" outcome, i.e., the outcome $\bar{\nu}$ such that $\bar{\nu}(\mathbf{1},\theta) = \mu(\theta)$ for all $\theta \in \Theta$ (which satisfies consistency by construction). The outcome $\bar{\nu}$ is *fully implementable in rationalizable strategies* if there exists an information structure in which the profile of the "all types play action 1" strategies is the unique interim correlated rationalizable strategy profile. An ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ satisfies *strong sequential obedience* if

$$\sum_{\gamma\in\Gamma_i,\theta\in\Theta} \nu_\Gamma(\gamma,\theta) \min_{a_{-i}\geq a_{-i}(\gamma)} d_i(a_{-i},\theta) > 0$$

for all $i \in I$ such that $\nu_\Gamma(\Gamma_i \times \Theta) > 0$; an outcome $\nu \in \Delta(A\times\Theta)$ satisfies strong sequential obedience if there exists an ordered outcome that induces ν and satisfies strong sequential obedience. Impose the dominance state assumption that there exists $\overline{\theta} \in \Theta$ such that for all $i \in I$, $d_i(a_{-i},\overline{\theta}) > 0$ for all $a_{-i} \in A_{-i}$.

Now consider the BAS game $(\underline{d}_i)_{i\in I}$ defined by $\underline{d}_i(a_{-i},\theta) = \min_{a'_{-i}\geq a_{-i}} d_i(a'_{-i},\theta)$ for all $i \in I$ and $a_{-i} \in A_{-i}$ (where the dominance state assumption is satisfied). Then, clearly, $\bar{\nu}$ is fully implementable in rationalizable strategies in $(d_i)_{i\in I}$ if and only if it is fully (hence S-)implementable in $(\underline{d}_i)_{i\in I}$, and ν_Γ satisfies strong sequential obedience in $(d_i)_{i\in I}$ if and only if it satisfies sequential obedience in $(\underline{d}_i)_{i\in I}$. Therefore, by Theorem 1, we have:

Proposition B.3. *Let $(d_i)_{i\in I}$ be any binary-action game. Then the outcome $\bar{\nu}$ is fully implementable in rationalizable strategies if and only if it satisfies strong sequential obedience.*

B.4.2. *Many Actions.* In this section, we consider games with many actions and present a generalized notion of sequential obedience that is necessary for S-implementability in these games. We also report a special case in which this notion is sufficient for S-implementability.

Let the action set A_i of player $i \in I$ be represented by a finite set of points in $[0,1]$ that contains 0 and 1 (so that $\min A_i = 0$ and $\max A_i = 1$), and let a base game $(u_i)_{i\in I}$ be given, where we assume supermodularity: for all $\theta \in \Theta$, $i \in I$, and $a_i, a'_i \in A_i$ with $a_i < a'_i$, $u_i((a'_i,a_{-i}),\theta) - u_i((a_i,a_{-i}),\theta)$ is nondecreasing in $a_{-i} \in A_{-i}$. The concept of smallest equilibrium implementability (S-implementability) is defined analogously to the case of binary actions.

A sequence $\gamma = (a^0, a^1, \dots, a^k)$ of action profiles is a *unilateral deviation path* from $\mathbf{0}$ if $a^0 = \mathbf{0}$, and for each $\ell = 1, \dots, k$, there exists $i_\ell \in I$ such that $a^\ell_{i_\ell} > a^{\ell-1}_{i_\ell}$ and $a^\ell_j = a^{\ell-1}_j$ for all $j \in I \setminus \{i_\ell\}$. Let Γ be the set of all unilateral deviation paths from $\mathbf{0}$. For $\gamma = (a^0, a^1, \dots, a^k) \in \Gamma$, denote $a(\gamma) = a^k$. An ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ induces an outcome $\nu \in \Delta(A \times \Theta)$ if

$$\nu(a, \theta) = \sum_{\gamma : a(\gamma) = a} \nu_\Gamma(\gamma, \theta).$$

For $i \in I$ and $a_i, a_i' \in A_i$ with $a_i < a_i'$, let $\Gamma_i(a_i, a_i') \subset \Gamma$ be the set of unilateral deviation paths along which player i switches from a_i to a_i', and for $\gamma \in \Gamma_i(a_i, a_i')$, let $a_{-i}(\gamma; a_i, a_i')$ be the profile of the opponents' actions when player i switches from a_i to a_i'.

An ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ satisfies sequential obedience if

$$\sum_{\gamma \in \Gamma_i(a_i, a_i'), \theta \in \Theta} \nu_\Gamma(\gamma, \theta)(u_i((a_i', a_{-i}(\gamma; a_i, a_i')), \theta) - u_i((a_i'', a_{-i}(\gamma; a_i, a_i')), \theta)) > 0 \qquad \text{(B.7)}$$

for all $i \in I$ and $a_i, a_i', a_i'' \in A_i$ with $a_i \le a_i'' < a_i'$ such that $\nu_\Gamma(\Gamma_i(a_i, a_i') \times \Theta) > 0$. An outcome $\nu \in \Delta(A \times \Theta)$ satisfies sequential obedience if there exists an ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ that induces ν and satisfies sequential obedience. If $|A_i| = 2$ for all $i \in I$, this coincides with the definition in the binary-action case.

Then, an argument almost identical with that in the proof of Theorem 1(1) shows that if an outcome is S-implementable, then it satisfies this version of sequential obedience along with consistency and obedience: in an implementing information structure, consider the sequential best response process from the smallest strategy and for any pair of actions $a_i < a_i'$, aggregate the obedience conditions upon the switch from a_i to a_i' in the process.

For sufficiency, we report a special case in which the generalized sequential obedience condition above, along with consistency, implies S-implementability. We focus on the "always play $\mathbf{1}$" outcome, i.e., the outcome $\bar{\nu}$ such that $\bar{\nu}(\mathbf{1}, \theta) = \mu(\theta)$ for all $\theta \in \Theta$. Let $\Pi \subset \Gamma$ be the set of unilateral deviation paths $\gamma = (a^0, a^1, \dots, a^{|I|})$ (of length $|I| + 1$) such that for each $\ell = 1, \dots, |I|$, $a^{\ell-1}_{i_\ell} = 0$ and $a^\ell_{i_\ell} = 1$ for some $i_\ell \in I$. Now assume the dominance state assumption, that there exists $\overline{\theta} \in \Theta$ such that for all $i \in I$, $u_i((1, \mathbf{0}_{-i}), \overline{\theta}) - u_i((a_i, \mathbf{0}_{-i}), \overline{\theta}) > 0$ for all $a_i < 1$, and suppose that the outcome $\bar{\nu}$ is induced by some ordered outcome $\nu_\Gamma \in \Delta(\Pi \times \Theta)$ that satisfies the sequential obedience condition (B.7), i.e., for all $i \in I$,

$$\sum_{\gamma \in \Pi, \theta \in \Theta} \nu_\Gamma(\gamma, \theta)(u_i((1, a_{-i}(\gamma; 0, 1)), \theta) - u_i((a_i, a_{-i}(\gamma; 0, 1)), \theta)) > 0 \qquad \text{(B.8)}$$

for all $a_i < 1$. Then, this condition can be thought of as $\nu_\Gamma \in \Delta(\Pi \times \Theta)$ satisfying sequential obedience in a game with *binary* actions. Formally, for each $(a_i)_{i\in I} \in \prod_{i\in I}(A_i \setminus \{1\})$, define $(d_i^{a_i})_{i\in I}$, $d_i^{a_i} \colon \{0,1\}^{I\setminus\{i\}} \times \Theta \to \mathbb{R}$, by

$$d_i^{a_i}(b_{-i},\theta) = u_i((1,b_{-i}),\theta) - u_i((a_i,b_{-i}),\theta).$$

Thus, the ordered outcome ν_Γ satisfies the condition (B.8) in $(u_i)_{i\in I}$ if and only if it satisfies sequential obedience in the BAS game $(d_i^{a_i})_{i\in I}$ for every $(a_i)_{i\in I} \in \prod_{i\in I}(A_i \setminus \{1\})$, where Π is naturally identified with the set of permutations of all players. Hence, the construction in the proof of Theorem 1(2) applies to this case, which shows that the outcome $\bar{\nu}$ is S-implementable in $(u_i)_{i\in I}$.

The condition discussed above is apparently very restrictive, but still broad enough to cover the result of Hoshino (2022) (which builds on the argument for Lemma 5.5 in Kajii and Morris (1997)). Assume that for every state $\theta \in \Theta$, action profile $\mathbf{1} \in A$ is a $\mathbf{p}(\theta)$-dominant equilibrium for some $\mathbf{p}(\theta) = (p_i(\theta))_{i\in I} \in [0,1]^I$ with $\sum_{i\in I} p_i(\theta) \leq 1$, i.e., for any $i \in I$ and any $a_i < 1$,

$$\sum_{a_{-i}\in A_{-i}} q_i(a_{-i})(u_i((1,a_{-i}),\theta) - u_i((a_i,a_{-i}),\theta)) \geq 0$$

for any $q_i \in \Delta(A_{-i})$ with $q_i(\mathbf{1}_{-i}) \geq p_i(\theta)$. Note that this condition is equivalently written as: for any $i \in I$ and any $a_i < 1$,

$$\sum_{b_{-i}\in\{0,1\}^{I\setminus\{i\}}} \underline{q}_i(b_{-i}) \min_{a_{-i}\geq b_{-i}} (u_i((1,a_{-i}),\theta) - u_i((a_i,a_{-i}),\theta)) \geq 0$$

for any $\underline{q}_i \in \Delta(\{0,1\}^{I\setminus\{i\}})$ with $\underline{q}_i(\mathbf{1}_{-i}) \geq p_i(\theta)$. Then, there exists some ordered outcome $\nu_\Gamma \in \Delta(\Pi \times \Theta)$ that induces $\bar{\nu}$ and satisfies the sequential obedience condition (B.8). To see this, for each $i \in I$, let $\gamma^i = (a^0, a^1, \ldots, a^{|I|}) \in \Pi$ be any path such that player i is the last player who switches (i.e., such that $a_i^{|I|-1} = 0$ and $a_i^{|I|} = 1$). Define $\nu_\Gamma^{\mathbf{p}(\cdot)} \in \Delta(\Pi \times \Theta)$ by

$$\nu_\Gamma^{\mathbf{p}(\cdot)}(\gamma,\theta) = \begin{cases} \dfrac{p_i(\theta)}{\sum_{j\in I} p_j(\theta)}\mu(\theta) & \text{if } \gamma = \gamma^i,\ i \in I, \\ 0 & \text{otherwise,} \end{cases}$$

which induces $\bar{\nu}$. Then (B.8) is satisfied: for any $i \in I$ and any $a_i < 1$,

$$\begin{aligned} &\sum_{\gamma\in\Pi,\theta\in\Theta} \nu_\Gamma^{\mathbf{p}(\cdot)}(\gamma,\theta)(u_i((1,a_{-i}(\gamma;0,1)),\theta) - u_i((a_i,a_{-i}(\gamma;0,1)),\theta)) \\ &= \sum_{\theta\in\Theta} \mu(\theta) \sum_{b_{-i}\in\{0,1\}^{I\setminus\{i\}}} \underline{q}_i^{\mathbf{p}(\theta)}(b_{-i})(u_i((1,b_{-i}),\theta) - u_i((a_i,b_{-i}),\theta)) > 0, \end{aligned}$$

where $\underline{q}_i^{\mathbf{P}(\theta)} \in \Delta(\{0,1\}^{I\setminus\{i\}})$ is defined by $\underline{q}_i^{\mathbf{P}(\theta)}(b_{-i}) = \frac{\nu_\Gamma^{\mathbf{P}(\cdot)}(\{\gamma\in\Pi | a_{-i}(\gamma;0,1)=b_{-i}\}\times\{\theta\})}{\mu(\theta)}$, which satisfies $\underline{q}_i^{\mathbf{P}(\theta)}(\mathbf{1}_{-i}) = \frac{p_i(\theta)}{\sum_{j\in I} p_j(\theta)} \geq p_i(\theta)$, so that $\sum_{b_{-i}\in\{0,1\}^{I\setminus\{i\}}} \underline{q}_i^{\mathbf{P}(\theta)}(b_{-i})(u_i((1,b_{-i}),\theta) - u_i((a_i,b_{-i}),\theta)) \geq 0$ holds for all $\theta \in \Theta$, with strict inequality for $\theta = \overline{\theta}$.

In fact, even if $(u_i)_{i\in I}$ is not supermodular, a stronger form of (B.8) holds, under the dominance state assumption that for all $i \in I$, $u_i((1,a_{-i}),\overline{\theta}) - u_i((a_i,a_{-i}),\overline{\theta}) > 0$ for all $a_i < 1$ and all $a_{-i} \in A_{-i}$:

$$\begin{aligned}&\sum_{\gamma\in\Pi,\theta\in\Theta} \nu_\Gamma^{\mathbf{P}(\cdot)}(\gamma,\theta) \min_{a_{-i}\geq a_{-i}(\gamma;0,1)} (u_i((1,a_{-i}),\theta) - u_i((a_i,a_{-i}),\theta))\\ &= \sum_{\theta\in\Theta} \mu(\theta) \sum_{b_{-i}\in\{0,1\}^{I\setminus\{i\}}} \underline{q}_i^{\mathbf{P}(\theta)}(b_{-i}) \min_{a_{-i}\geq b_{-i}} (u_i((1,a_{-i}),\theta) - u_i((a_i,a_{-i}),\theta)) > 0.\end{aligned}$$

Therefore, as argued in Section B.4.1, this implies that the outcome $\bar{\nu}$ is fully implementable in rationalizable strategies with or without supermodularity, which reproduces the result of Hoshino (2022, Theorem 1).

B.4.3. *Adversarial Information Sharing.* In this section, we formulate and study an optimal information design problem where the designer is also concerned that information might be shared among players in an adversarial way.

For an information structure $\mathcal{T} = ((T_i)_{i\in I}, \pi)$, an information structure $\mathcal{T}' = ((T_i')_{i\in I}, \pi')$ is an *information sharing* of $\mathcal{T}$ if there exist a profile $(Z_i)_{i\in I}$ of sets of "supplementary signals" and a signal generation rule $\phi\colon T \to \Delta(Z)$ such that $T_i' = T_i \times Z_i$ for each $i \in I$, and $\pi'(t,z,\theta) = \pi(t,\theta)\phi(t)(z)$ for any $t \in T$, $z \in Z$, and $\theta \in \Theta$. The value of information design under adversarial information sharing (and adversarial equilibrium selection) is then formulated as

$$V^\dagger = \sup_{\mathcal{T}} \inf_{\mathcal{T}'\text{: information sharing of }\mathcal{T}} \min_{\sigma\in E(\mathcal{T}')} \sum_{t'\in T',\theta\in\Theta} \pi'(t',\theta)V(\sigma(t'),\theta). \tag{B.9}$$

We first develop a useful alternative representation of this problem. A strategy profile $\sigma' = (\sigma_i')_{i\in I}$, $\sigma_i'\colon T_i \times Z_i \to \Delta(A_i)$, in the information sharing $\mathcal{T}' = ((T_i \times Z_i)_{i\in I}, \pi')$ of $\mathcal{T}$ induces an outcome $\xi \in \Delta(A \times T)$:

$$\begin{aligned}\xi(a,t) &= \sum_{z\in Z,\theta\in\Theta} \pi'(t,z,\theta) \prod_{i\in I} \sigma_i'(t_i,z_i)(a_i)\\ &= \pi(t) \sum_{z\in Z} \phi(t)(z) \prod_{i\in I} \sigma_i'(t_i,z_i)(a_i),\end{aligned}$$

where $\pi(t) = \sum_{\theta\in\Theta} \pi(t,\theta)$. Thus, the value of V under σ' is written as

$$\sum_{t'\in T',\theta\in\Theta} \pi'(t',\theta)V(\sigma'(t'),\theta) = \sum_{a\in A,\theta\in\Theta} \xi(a,t)V^{\mathcal{T}}(a,t),$$

where $V^{\mathcal{T}}(a,t) = \sum_{\theta\in\Theta} \pi(\theta|t)V(a,\theta)$ with $\pi(\theta|t) = \frac{\pi(t,\theta)}{\pi(t)}$. Say that an outcome $\xi \in \Delta(A \times T)$ is a *Bayesian solution* (Forges (1993)) of $\mathcal{T}$ if it satisfies consistency for $\mathcal{T}$: $\sum_{a\in A} \xi(a,t) = \pi(t)$ for all $t \in T$, and obedience for $\mathcal{T}$:

$$\sum_{a_{-i}\in A_{-i},t_{-i}\in T_{-i}} \xi((a_i,a_{-i}),(t_i,t_{-i})) \left(u_i^{\mathcal{T}}((a_i,a_{-i}),(t_i,t_{-i})) - u_i^{\mathcal{T}}((a_i',a_{-i}),(t_i,t_{-i}))\right) \geq 0$$

for all $i \in I$, $t_i \in T_i$, and $a_i, a_i' \in A_i$, where $u_i^{\mathcal{T}}(a,t) = \sum_{\theta\in\Theta} \pi(\theta|t)u_i(a,\theta)$. One can show that an outcome $\xi \in \Delta(A\times T)$ is induced by some equilibrium of some information sharing of $\mathcal{T}$ if and only if it is a Bayesian solution of $\mathcal{T}$ (see Bergemann and Morris (2016)). Let $BS(\mathcal{T}) \subset \Delta(A \times T)$ denote the set of Bayesian solutions of $\mathcal{T}$. Thus, the original problem (B.9) is rewritten as

$$V^\dagger = \sup_{\mathcal{T}} \min_{\xi\in BS(\mathcal{T})} \sum_{a\in A,t\in T} \xi(a,t)V^{\mathcal{T}}(a,t). \tag{B.10}$$

We want to compare this problem with the constrained problem where the designer can only send public information to the players. We say that an information structure $\mathcal{T} = ((T_i)_{i\in I},\pi)$ is *public* if $T_i = T_j$ for all $i,j \in I$, and $\sum_{t\in T:t_1=\cdots=t_{|I|},\theta\in\Theta} \pi(t,\theta) = 1$. Under the supermodularity of the payoffs and the monotonicity of V, the value of information design under S-implementation (thus the superscript "S") with public information structures is

$$V^{\mathrm{S}}_{\mathrm{public}} = \sup_{\mathcal{T}:\,\mathrm{public}} \min_{\sigma\in E(\mathcal{T})} \sum_{t\in T,\theta\in\Theta} \pi(t,\theta)V(\sigma(t),\theta) \tag{B.11}$$

$$= \sup_{\mathcal{T}:\,\mathrm{public}} \sum_{t\in T} \pi(t)V^{\mathcal{T}}(\underline{\sigma}(\mathcal{T})(t),t). \tag{B.12}$$

We immediately have $V^\dagger \leq V^{\mathrm{S}}_{\mathrm{public}}$ (compare the expressions (B.9) and (B.11)). On the other hand, if $\mathcal{T}$ is a public information structure, then for any $\xi \in BS(\mathcal{T})$ and for any $t \in T$, $\xi(\cdot|t) = \frac{\xi(\cdot,t)}{\pi(t)} \in \Delta(A)$ is a correlated equilibrium of $(u_i^{\mathcal{T}}(\cdot,t))_{i\in I}$, and by supermodularity, $BS(\mathcal{T})$ has a smallest element, which equals $\underline{\sigma}(\mathcal{T})$; thus we have $V^\dagger \geq V^{\mathrm{S}}_{\mathrm{public}}$ (compare (B.10) and (B.12)). Hence, by supermodularity and the objective monotonicity, we have:

Proposition B.4. $V^\dagger = V^{\mathrm{S}}_{\mathrm{public}}$.

Note that this result holds with any number of actions.

The set of outcomes that are S-implementable by public information structures can be characterized by the following strengthening of sequential obedience. Say that an ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ satisfies *public sequential obedience* (resp. *weak public sequential* *obedience*) if for every $\gamma \in \Gamma$ such that $\sum_{\theta\in\Theta} \nu_\Gamma(\gamma, \theta) > 0$,

$$\sum_{\theta\in\Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > \text{(resp. } \geq \text{) } 0 \tag{B.13}$$

for all $i \in S(\gamma)$. An outcome $\nu \in \Delta(A \times \Theta)$ satisfies public sequential obedience (resp. weak public sequential obedience) if there exists an ordered outcome that induces ν and satisfies public sequential obedience (resp. weak public sequential obedience). It can readily be shown that an outcome is S-implementable by a public information structure if and only if it satisfies consistency, obedience, and public sequential obedience. Therefore, the problem (B.12) (hence (B.9)) can also be written as

$$V^{\text{S}}_{\text{public}} = \max_{\nu\in\Delta(A\times\Theta)} \sum_{a\in A, \theta\in\Theta} \nu(a, \theta) V(a, \theta)$$

subject to consistency and weak public sequential obedience. This also has a concavification form:

$$V^{\text{S}}_{\text{public}} = \sup_{Q\in\Delta_0(\Delta(\Theta))} \sum_{q\in\operatorname{supp} Q} Q(q) \sum_{\theta\in\Theta} q(\theta) V(\underline{a}(q), \theta)$$

subject to

$$\sum_{q\in\operatorname{supp} Q} Q(q) q(\theta) = \mu(\theta) \text{ for all } \theta \in \Theta,$$

where $\Delta_0(\Delta(\Theta))$ denotes the set of distributions over $\Delta(\Theta)$ with finite support (with $\operatorname{supp} Q$ denoting the support of $Q \in \Delta_0(\Delta(\Theta))$), and $\underline{a}(q) \in A$ denotes the smallest Nash equilibrium of the average game given by the common posterior $q \in \Delta(\Theta)$: $\sum_{\theta\in\Theta} q(\theta) u_i(a, \theta)$.

Let $V^{\text{S}}_{\text{private}}$ denote the optimal value of the unconstrained problem (i.e., the problem under S-implementation with general information structures). Then $V^{\text{S}}_{\text{private}} \geq V^{\text{S}}_{\text{public}}$ trivially, and a strict inequality holds—or equivalently, the unconstrained optimal outcome is not S-implementable by a public information structure—for example under the assumptions in Section 4 if in addition there is nontrivial strategic interdependence among players in that $\Phi((1, \mathbf{0}_{-i}), \theta) < \frac{1}{|I|} \Phi(\mathbf{1}, \theta)$ for some $i \in I$ and some $\theta > \theta^*$. To see this, let ν^* be the optimal (perfect coordination) outcome given in (4.2) and let Π^i be the set of permutations of all players in which i appears first. Then for any ordered outcome ν_Γ that induces ν^*, we have

$$\begin{aligned}\sum_{i\in I, \gamma\in\Pi^i} \sum_{\theta\in\Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) &< \sum_{i\in I, \gamma\in\Pi^i} \sum_{\theta\in\Theta} \nu_\Gamma(\gamma, \theta) \frac{1}{|I|} \Phi(\mathbf{1}, \theta) \\ &= \frac{1}{|I|} \sum_{\theta\in\Theta} \nu^*(\mathbf{1}, \theta) \Phi(\mathbf{1}, \theta) = 0,\end{aligned}$$

so that weak public sequential obedience is violated for some $i \in I$ and $\gamma \in \Pi^i$.

In Section B.1, we illustrated the solutions under public information structures as well as under private (i.e., general) information structures for the example in Section 2.

B.4.4. *Finite Information Structures.* The construction in the proof of Theorem 1(2) involves infinitely many types, but a similar construction with a finite number of types with a uniform—instead of geometric—distribution for the variable m can be used to S-implement the same outcome, where the number of types will be large enough depending on the probability $\mu(\overline{\theta})$ of the dominance state $\overline{\theta}$ and the degree of dominance at $\overline{\theta}$ relative to the payoffs at other states as well as the slackness of sequential obedience of the given outcome to be implemented. Specialized to a symmetric two-player two-state example, Mathevet et al. (2020) present an information structure with three types or less for each player that implements the optimal outcome. Our example in Section 2 also allows such a small information structure. Again specialized to a particular class of games (i.e., regime change games), Li et al. (2022) identify the unique optimal information structure when the number of types of each player is constrained by some upper bound K and show that their unconstrained optimal information structure is obtained as the limit of those finite information structures as $K \to \infty$.

While we do not need literally infinitely many types as argued above, our results rely on the assumption that there is no a priori bound on the number of types. To see what would happen otherwise, suppose that $\mathbf{0}$ is a strict equilibrium at every state $\theta \neq \overline{\theta}$. Let $SI^K(\mu)$ denote the set of outcomes that are S-implementable, under prior μ, by some information structure such that the number of types of each player is at most K. By Kajii and Morris (1997, Lemmas 5.2 and B), we immediately have the following.

Proposition B.5. *Suppose that* $\mathbf{0}$ *is a strict equilibrium at every* $\theta \neq \overline{\theta}$*. Then for any* $K < \infty$ *and* $\delta > 0$*, there exists* $\varepsilon > 0$ *such that if* $\mu(\overline{\theta}) \leq \varepsilon$*, then for any* $\nu \in SI^K(\mu)$*, we have* $\sum_{\theta \in \Theta} \nu(\mathbf{0}, \theta) \geq 1 - \delta$*.*

That is, if there is a bound on the number of types, then as the probability of the dominance state vanishes, the S-implementable outcomes tend only to be the trivial outcome “always play $\mathbf{0}$”. Note that this result holds for general games (with any finite number of actions and possibly non-supermodular payoffs).

B.4.5. *Uncountable Information Structures.* In this section, we demonstrate that Theorem 1(1) continues to hold with possibly uncountable type spaces.

Let the finite state space Θ with prior $\mu \in \Delta(\Theta)$ and the base game $(d_i)_{i\in I}$ be given as in Section 1.2. An information structure is defined as follows. For each player $i \in I$, the set T_i of types is a measurable space endowed with sigma-algebra $\mathcal{F}_i$, where we write $T = \prod_{i\in I} T_i$ and $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i$. The common prior π is a probability measure on $T \times \Theta$ (endowed with the product sigma-algebra $\mathcal{F} \otimes 2^{\Theta}$). Let π_X denote the marginal of π on $X = T_i, T, \Theta$, etc. We require π to be consistent with μ, i.e., $\pi_\Theta = \mu$.

In the incomplete information game induced by an information structure $((T_i, \mathcal{F}_i)_{i\in I}, \pi)$, a (pure) strategy for player $i \in I$ is an equivalence class of measurable functions from T_i to A_i modulo being equal π_{T_i}-a.s. Let Σ_i be the partially ordered set of strategies of i, where $\sigma_i \leq \sigma'_i$ is understood as $\sigma_i(t_i) \leq \sigma'_i(t_i)$ for π_{T_i}-a.s. $t_i \in T_i$. Write $\Sigma = \prod_{i\in I} \Sigma_i$ and $\Sigma_{-i} = \prod_{j\neq i} \Sigma_j$ (endowed with the product partial orders, respectively). For $i \in I$, $t_i \in T_i$, and $\sigma_{-i} \in \Sigma_{-i}$, write

$$D_i(\sigma_{-i}|t_i) = \mathbb{E}[d_i(\sigma_{-i}(\cdot), \cdot)|\mathcal{F}_i](t_i),$$

which is measurable in t_i and nondecreasing in σ_{-i}, and let $\beta_i(\sigma_{-i})$ be the set of best responses to σ_{-i}, i.e., the set of strategies $\sigma_i \in \Sigma_i$ such that for π_{T_i}-a.s. $t_i \in T_i$, $D_i(\sigma_{-i}|t_i) \geq 0$ (resp. $D_i(\sigma_{-i}|t_i) \leq 0$) if $\sigma_i(t_i) = 1$ (resp. $\sigma_i(t_i) = 0$). A strategy profile $\sigma = (\sigma_i)_{i\in I} \in \Sigma$ is an equilibrium if for all $i \in I$, $\sigma_i \in \beta_i(\sigma_{-i})$. By the supermodularity (and the boundedness and the continuity of $d_i(a_{-i}, \theta)$ in a_{-i}), a smallest equilibrium exists and it is the limit of sequential best responses from the smallest strategy profile, as we now show below.

For $i \in I$, define $\underline{\beta}_i \colon \Sigma_{-i} \to \Sigma_i$ by

$$\underline{\beta}_i(\sigma_{-i})(t_i) = \begin{cases} 1 & \text{if } D_i(\sigma_{-i}|t_i) > 0, \\ 0 & \text{if } D_i(\sigma_{-i}|t_i) \leq 0 \end{cases}$$

for π_{T_i}-a.s. $t_i \in T_i$, which is well defined (measurable in t_i and unique up to π_{T_i}-a.s.) and nondecreasing in σ_{-i}. By construction, $\underline{\beta}_i(\sigma_{-i})$ is the smallest element of $\beta_i(\sigma_{-i})$. Then define the sequence of strategy profiles $\{\sigma^n\}$ as follows: let $\sigma_i^0(t_i) = 0$ for all $i \in I$ and $t_i \in T_i$, and for $n = 1, 2, \ldots$, let

$$\sigma_i^n = \begin{cases} \underline{\beta}_i(\sigma_{-i}^{n-1}) & \text{if } i \equiv n \pmod{|I|}, \\ \sigma_i^{n-1} & \text{otherwise.} \end{cases}$$

By the monotonicity of $\underline{\beta}_i$, this sequence is monotone increasing, $\sigma^0 \leq \sigma^1 \leq \cdots$, and converges as $n \to \infty$ to some $\underline{\sigma} \in \Sigma$ π_T-a.s. Since for each $i \in I$, $D_i(\underline{\sigma}_{-i}|t_i) =$

$\lim_{n\to\infty} D_i(\sigma^n_{-i}|t_i)$ for π_{T_i}-a.s. by the dominated convergence theorem, $\underline{\sigma}$ is an equilibrium, and again by the monotonicity of $\underline{\beta}_i$, it is the smallest equilibrium.

Now we show the necessity of sequential obedience for S-implementability within this framework. Let $\nu \in \Delta(A \times \Theta)$ be S-implementable, and $((T_i, \mathcal{F}_i)_{i\in I}, \pi)$ be an information structure whose smallest equilibrium $\underline{\sigma}$ induces ν, i.e., $\nu(a, \theta) = \pi(\{t \in T \mid \underline{\sigma}(t) = a\} \times \{\theta\})$. Define the sequence of strategy profiles $\{\sigma^n\}$ as above, and let $\bar{T}_i \in \mathcal{F}_i$ be the set of types such that $\{\sigma^n_i(t_i)\}$ is monotone, where $\pi_{T_i}(\bar{T}_i) = 1$. On $\bar{T} = \prod_{i\in I} \bar{T}_i$, define $n_i(t_i)$ and $T(\gamma)$, $\gamma \in \Gamma$, as in the proof of Theorem 1(1), where one can verify that $T(\gamma) \in \mathcal{F}$. Then define $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ by $\nu_\Gamma(\gamma, \theta) = \pi(T(\gamma) \times \{\theta\})$, which induces ν. We want to show that ν_Γ satisfies sequential obedience.

Fix any $i \in I$ with $\nu_\Gamma(\Gamma_i \times \Theta) > 0$, where $\pi_{T_i}(\{t_i \in \bar{T}_i \mid n_i(t_i) = n\}) > 0$ for some $n \in \mathbb{N}$. Note that for all $n \in \mathbb{N}$ and all $t_i \in \bar{T}_i$ with $n_i(t_i) = n$, we have $D_i(\sigma^{n-1}_{-i}|t_i) > 0$. Hence, we have

$$
\begin{aligned}
0 &< \sum_{n\in\mathbb{N}} \int_{\{t_i \in \bar{T}_i \mid n_i(t_i)=n\}} D_i(\sigma^{n-1}_{-i}|t_i) d\pi_{T_i}(t_i) \\
&= \sum_{n\in\mathbb{N}} \int_{\{t_i \in \bar{T}_i \mid n_i(t_i)=n\}\times \bar{T}_{-i}\times\Theta} d_i(\sigma^{n-1}_{-i}(t_{-i}), \theta) d\pi(t, \theta) \\
&= \sum_{n\in\mathbb{N}} \sum_{a_{-i}\in A_{-i}, \theta\in\Theta} d_i(a_{-i}, \theta)\pi(\{t \in \bar{T} \mid n_i(t_i) = n,\ \sigma^{n-1}_{-i}(t_{-i}) = a_{-i}\} \times \{\theta\}) \\
&= \sum_{a_{-i}\in A_{-i}, \theta\in\Theta} d_i(a_{-i}, \theta)\pi(\{t \in \bar{T} \mid n_i(t_i) < \infty,\ \sigma^{n_i(t_i)-1}_{-i}(t_{-i}) = a_{-i}\} \times \{\theta\}) \\
&= \sum_{a_{-i}\in A_{-i}, \theta\in\Theta} d_i(a_{-i}, \theta) \sum_{\gamma\in\Gamma_i : a_{-i}(\gamma)=a_{-i}} \pi(T(\gamma) \times \{\theta\}) \\
&= \sum_{\gamma\in\Gamma_i, \theta\in\Theta} d_i(a_{-i}(\gamma), \theta)\pi(T(\gamma) \times \{\theta\}) = \sum_{\gamma\in\Gamma_i, \theta\in\Theta} d_i(a_{-i}(\gamma), \theta)\nu_\Gamma(\gamma, \theta),
\end{aligned}
$$

as desired.

B.4.6. *Dominance State Assumption.* The dominance state assumption, that there exists $\overline{\theta} \in \Theta$ such that $d_i(\mathbf{0}_{-i}, \overline{\theta}) > 0$ for all $i \in I$, is maintained throughout the analysis and used in Theorem 1(2) (and other results that use Theorem 1(2)). This exact form of the assumption, however, is stronger than needed and can be relaxed. Consider instead the following weakening, say the "sequential dominance states assumption": there exist a permutation γ of all players and states $\overline{\theta}^i \in \Theta$, $i \in I$, such that $d_i(a_{-i}(\gamma), \overline{\theta}^i) > 0$ for all $i \in I$. Under this condition, if ν satisfies consistency, obedience, sequential obedience, and grain of dominance with respect to $(\overline{\theta}^i)_{i\in I}$, i.e., $\nu(\mathbf{1}, \overline{\theta}^i) > 0$ for each $i \in I$, then one can construct an information structure, similar to, but more involved than, the one in the proof of Theorem 1(2), that S-implements ν. Thus, we have:

Proposition B.6. *If the sequential dominance states assumption is satisfied with respect to $(\overline{\theta}^i)_{i\in I}$, and ν satisfies consistency, obedience, sequential obedience, and grain of dominance with respect to $(\overline{\theta}^i)_{i\in I}$, then $\nu \in SI$.*

Conversely, if there exists $\nu \in SI$ such that $\nu(\mathbf{1},\theta) > 0$ for some $\theta \in \Theta$, then the sequential dominance states assumption must be satisfied.[33] Thus, it is weakest possible in this sense.

B.4.7. *Indispensability of Grain of Dominance.* This section presents an example demonstrating that the grain of dominance property is indispensable in Theorem 1(2).

Consider the following game: Let $I = \{1, 2\}$ and $\Theta = \{\theta_1, \theta_2, \theta_3\}$, and let $\mu(\theta) = \frac{1}{3}$ for all $\theta \in \Theta$. The payoffs for each $i \in I$ are given by

$$d_i(a_j, \theta_1) = -2,\ d_i(a_j, \theta_3) = 1$$

for all $a_j \in A_j$, $j \neq i$, and

$$d_i(0, \theta_2) = -1,\ d_i(1, \theta_2) = 2.$$

The dominance state assumption is satisfied with $\overline{\theta} = \theta_3$.

Let $\nu \in \Delta(A \times \Theta)$ be defined by $\nu(\mathbf{0}, \theta) = \mu(\theta)$ for $\theta = \theta_1, \theta_3$ and $\nu(\mathbf{1}, \theta_2) = \mu(\theta_2)$ (and $\nu(a, \theta) = 0$ otherwise). It satisfies consistency, obedience, and sequential obedience (for example with $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ such that $\nu_\Gamma(\emptyset, \theta) = \nu(\mathbf{0}, \theta)$ for $\theta = \theta_1, \theta_3$ and $\nu_\Gamma(12, \theta_2) = \nu_\Gamma(21, \theta_2) = \frac{\nu(\mathbf{1},\theta_2)}{2}$), but not grain of dominance. We claim that $\nu \notin SI$.

Let $\mathcal{T} = ((T_i)_{i\in I}, \pi)$ be any information structure that has an equilibrium σ that induces ν. We show that the smallest strategy profile σ^0, the strategy profile such that $\sigma_i^0(0|t_i) = 1$ for all t_i, is an equilibrium (hence the smallest equilibrium) in any such $\mathcal{T}$. Let

$$T_i^{a_i} = \{t_i \in T_i \mid \sigma_i(a_i|t_i) = 1\},$$

and $T^a = T_1^{a_1} \times T_2^{a_2}$. By the assumption that σ induces ν, we have

$$\pi(T^{\mathbf{1}} \times \{\theta_2\}) = \nu(\mathbf{1}, \theta_2) = \mu(\theta_2) = \pi(T \times \{\theta_2\}),$$

[33] If such an S-implementable outcome exists, then for any $S \subsetneq I$, there exist $i \in I \setminus S$ and $\theta \in \Theta$ such that $d_i(\mathbf{1}_S, \theta) > 0$ (if there is no such pair of $i \in I \setminus S$ and $\theta \in \Theta$, then for any $\nu \in SI$, we would have $\nu(a, \theta) = 0$ for all $\theta \in \Theta$ whenever $a_i = 1$ for some $I \setminus S$). Then inductively apply this condition to construct γ and $\overline{\theta}^i$ in the sequential dominance state assumption.

and hence $\pi(\theta_2|t_i) = 1$ for all $t_i \in T_i^1$. Therefore, for all $t_i \in T_i^1$, $a_i = 0$ is a best response against σ_j^0. For $t_i \in T_i^0$, $a_i = 0$, which is a best response against σ_j, continues to be a best response against σ_j^0 by supermodularity. This shows that σ^0 is the smallest equilibrium for any information structure that partially implements ν, which implies that $\nu \notin SI$.

B.5. **Perfect Coordination in Generalized Regime Change Games.** In this section, we show that the perfect coordination property holds—i.e., an optimal outcome is found among perfect coordination outcomes—in generalized regime change games.

The base game $(d_i)_{i\in I}$ is a *generalized regime change game* if there exists a function $r\colon A \times \Theta = \{0,1\}$ such that for all $i \in I$,

$$d_i(a_{-i},\theta)\begin{cases} > 0 & \text{if } r((1,a_{-i}),\theta) = 1, \\ \leq 0 & \text{if } r((1,a_{-i}),\theta) = 0. \end{cases}$$

By supermodularity, $r(a,\theta)$ must be nondecreasing in a. To be consistent with the dominance state assumption, we assume that for some $\overline{\theta} \in \Theta$, $r((1,\mathbf{0}_{-i}),\overline{\theta}) = 1$ for all $i \in I$. This game corresponds to that considered in Inostroza and Pavan (2022, Additional Material, Section AM3). The following proposition is a version of their Theorem AM3-1 restricted to our setting, where our concise proof appeals to our sequential obedience characterization of Theorem 1.

Proposition B.7. *Let $(d_i)_{i\in I}$ be a generalized regime change game. For any outcome $\nu \in \overline{SI}$, there exists a perfect coordination outcome $\hat{\nu} \in \overline{SI}$ such that $\sum_{a\in A:r(a,\theta)=1}\hat{\nu}(a,\theta) = \sum_{a\in A:r(a,\theta)=1}\nu(a,\theta)$ for all $\theta \in \Theta$.*

Proof. Let $(d_i)_{i\in I}$ be a generalized regime change game, and let $\nu \in \overline{SI}$. By Corollary 1 to Theorem 1, ν satisfies consistency, obedience, and weak sequential obedience. By Proposition 2, it satisfies condition (3.8). Define $\hat{\nu} \in \Delta(A \times \Theta)$ by

$$\hat{\nu}(a,\theta) = \begin{cases} \sum_{a'\in A:r(a',\theta)=1}\nu(a',\theta) & \text{if } a = \mathbf{1}, \\ \sum_{a'\in A:r(a',\theta)=0}\nu(a',\theta) & \text{if } a = \mathbf{0}, \\ 0 & \text{otherwise}, \end{cases}$$

which by construction satisfies consistency and perfect coordination. It also satisfies, for each $\theta \in \Theta$, $\sum_{a\in A:r(a,\theta)=1}\nu(a,\theta) = \sum_{a\in A:r(a,\theta)=1}\hat{\nu}(a,\theta)$ since by the monotonicity of $r(a,\theta)$ in a, $\{a \in A \mid r(a,\theta) = 1\} \neq \emptyset$ if and only if $r(\mathbf{1},\theta) = 1$. We want to show that $\hat{\nu}$ satisfies weak sequential obedience and lower obedience.

First, for weak sequential obedience, we show that $\hat{\nu}$ satisfies condition (3.8) in Proposition 2. Indeed, for any $(\lambda_i)_{i\in I} \in \mathbb{R}^I_+$, we have

$$\begin{aligned}
&\sum_{a\in A,\theta\in\Theta} \hat{\nu}(a,\theta) \max_{\gamma:a(\gamma)=a} \sum_{i\in S(a)} \lambda_i d_i(a_{-i}(\gamma),\theta) \\
&\qquad = \sum_{\theta\in\Theta} \hat{\nu}(\mathbf{1},\theta) \max_{\gamma:a(\gamma)=\mathbf{1}} \sum_{i\in I} \lambda_i d_i(a_{-i}(\gamma),\theta) \\
&\qquad = \sum_{\substack{a\in A,\theta\in\Theta \\ r(a,\theta)=1}} \nu(a,\theta) \max_{\gamma:a(\gamma)=\mathbf{1}} \sum_{i\in I} \lambda_i d_i(a_{-i}(\gamma),\theta) \\
&\qquad \geq \sum_{\substack{a\in A,\theta\in\Theta \\ r(a,\theta)=1}} \nu(a,\theta) \max_{\gamma:a(\gamma)=a} \sum_{i\in I} \lambda_i d_i(a_{-i}(\gamma),\theta) \\
&\qquad \geq \sum_{\substack{a\in A,\theta\in\Theta \\ r(a,\theta)=1}} \nu(a,\theta) \max_{\gamma:a(\gamma)=a} \sum_{i\in S(a)} \lambda_i d_i(a_{-i}(\gamma),\theta) \\
&\qquad \geq \sum_{a\in A,\theta\in\Theta} \nu(a,\theta) \max_{\gamma:a(\gamma)=a} \sum_{i\in S(a)} \lambda_i d_i(a_{-i}(\gamma),\theta) \geq 0,
\end{aligned}$$

where the first inequality holds by supermodularity (for $i \notin S(\gamma)$, $a_{-i}(\gamma)$ denotes the action profile of i's opponents such that player j plays action 1 if and only if $j \in S(\gamma)$), the second inequality holds since $d_i(a_{-i},\theta) > 0$ when $r((0,a_{-i}),\theta) = 1$, the third inequality holds since $d_i(a_{-i},\theta) \leq 0$ when $r((1,a_{-i}),\theta) = 0$, and the last inequality holds by condition (3.8) for ν_Γ. Therefore, by Proposition 2, $\hat{\nu}$ satisfies weak sequential obedience.

Second, for lower obedience (i.e., condition (1.1) with $a_i = 0$), for each $i \in I$ we have

$$\begin{aligned}
&\sum_{a_{-i}\in A_{-i},\theta\in\Theta} \hat{\nu}((0,a_{-i}),\theta) d_i(a_{-i},\theta) = \sum_{\theta\in\Theta} \hat{\nu}(\mathbf{0},\theta) d_i(\mathbf{0}_{-i},\theta) = \sum_{\substack{a\in A,\theta\in\Theta: \\ r(a,\theta)=0}} \nu(a,\theta) d_i(\mathbf{0}_{-i},\theta) \\
&\leq \sum_{\substack{a\in A,\theta\in\Theta: \\ r(a,\theta)=0}} \nu(a,\theta) d_i(a_{-i},\theta) \\
&= \sum_{\substack{a_{-i}\in A_{-i},\theta\in\Theta: \\ r((1,a_{-i}),\theta)=0}} \nu((1,a_{-i}),\theta) d_i(a_{-i},\theta) + \sum_{\substack{a_{-i}\in A_{-i},\theta\in\Theta: \\ r((0,a_{-i}),\theta)=0}} \nu((0,a_{-i}),\theta) d_i(a_{-i},\theta) \\
&\leq \sum_{\substack{a_{-i}\in A_{-i},\theta\in\Theta: \\ r((0,a_{-i}),\theta)=0}} \nu((0,a_{-i}),\theta) d_i(a_{-i},\theta) \\
&\leq \sum_{a_{-i}\in A_{-i},\theta\in\Theta} \nu((0,a_{-i}),\theta) d_i(a_{-i},\theta) \leq 0,
\end{aligned}$$

where the first inequality holds by supermodularity, the second inequality holds since $d_i(a_{-i},\theta) \leq 0$ when $r((1,a_{-i}),\theta) = 0$, the third inequality holds since $d_i(a_{-i},\theta) > 0$ when

$r((0, a_{-i}), \theta) = 1$, and the last inequality holds by the lower obedience of ν. Therefore, $\hat{\nu}$ satisfies lower obedience.

Hence, we have $\hat{\nu} \in \overline{SI}$ by Corollary 1. □

For a generalized regime change game $(d_i)_{i \in I}$, the objective V is a *generalized regime change objective* with respect to $(d_i)_{i \in I}$ if it is written as

$$V(a, \theta) = \begin{cases} > 0 & \text{if } r(a, \theta) = 1, \\ = 0 & \text{if } r(a, \theta) = 0, \end{cases}$$

where we maintain the assumption that $V(a, \theta)$ is nondecreasing in a. By Proposition B.7, we immediately have the following.

Proposition B.8. *Let $(d_i)_{i \in I}$ be a generalized regime change game, and V a generalized regime change objective with respect to $(d_i)_{i \in I}$. Then there exists an optimal outcome of the adversarial information design problem that satisfies perfect coordination.*

Like Inostroza and Pavan (2022), we are not able to obtain explicitly the solution to the problem at this level of generality. Under the assumption of the existence of a convex potential (which covers symmetric regime change games), we derived an explicit expression of the optimal perfect coordination outcome in Theorem 2 in Section 4.