Econometrica, Vol. 92, No. 3 (May, 2024), 775–813

IMPLEMENTATION VIA INFORMATION DESIGN IN BINARY-ACTION SUPERMODULAR GAMES

STEPHEN MORRIS
Department of Economics, Massachusetts Institute of Technology

DAISUKE OYAMA
Faculty of Economics, University of Tokyo

SATORU TAKAHASHI
Faculty of Economics, University of Tokyo and Department of Economics, National University of Singapore

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 profile played on the largest event where that action profile maximizes the average potential.

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

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*.

Stephen Morris: semorris@mit.edu
Daisuke Oyama: oyama@e.u-tokyo.ac.jp
Satoru Takahashi: satorut@e.u-tokyo.ac.jp

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, 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, Bar Ilan, and California Los Angeles Universities. Stephen Morris gratefully acknowledges financial support from NSF Grants 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.

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, that is, 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 (Bergemann and Morris (2016)).

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*, where each player has two actions, low and high, and the payoffs are supermodular at each state. Smallest equilibrium implementation requires that the outcome be induced in the smallest equilibrium of the incomplete information game defined by the chosen information structure.[1] A smallest equilibrium always exists, and it arises if players are initially playing the low action and switch to the high action only if it is uniquely rationalizable to do so.[2] Our first main result addresses the implementability question by providing a characterization of smallest equilibrium implementability.

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 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 (there exists a state at which the high action is a dominant action for all players), we show that (the closure of) the set of smallest equilibrium implementable outcomes is equal to the set of outcomes (consistent with the prior) that satisfy sequential obedience as well as obedience.[3] Thus, this set, 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, that is, smallest equilibrium implementable outcomes, having been characterized by our first result, our second result determines which outcomes in this set

[1]Strategy profiles are partially ordered according to the probability of playing the high action at each type of each player.

[2]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.

[3]Largest equilibrium implementability can symmetrically be characterized by a reverse version of sequential obedience. For full implementability, that is, the requirement that an outcome be induced by all equilibria of some information structure, sequential obedience and reverse sequential obedience are clearly necessary, and we show in Appendix B.1 of the Supplemental Material (Morris, Oyama, and Takahashi (2024)) that, under an appropriate extension of the dominance state assumption, these conditions are in fact jointly sufficient.

[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.

are optimal given the objective function of the designer. This result applies when two restrictions are satisfied. First, the game has a potential (Monderer and Shapley (1996));[5] this 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. Second, the potential and designer's objective are convex;[6] this restriction is automatically satisfied in symmetric games (by supermodularity) and is satisfied in asymmetric games when the asymmetry is 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, that is, her intrinsic preference for coordination, only increases the advantages of perfect coordination. Solving our information design problem then reduces to solving a simple 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.

In the recent literature on information design with adversarial equilibrium selection, the problem has been addressed only in some special BAS games. The present paper offers a unified explanation, for general BAS games. It also provides 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. We discuss the related literature in Section 5, as well as in Section 2 through the leading example and in Section 3.3 through the (limit) complete information case.

1.1. *Setting*

There are finitely many players, denoted by $I = \{1, \ldots, |I|\}$, $|I| \geq 2$. 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*, that is, for each $i \in I$ and $\theta \in \Theta$,

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

is weakly increasing in $a_{-i} \in A_{-i}$. We denote $\mathbf{0} = (0, \ldots, 0) \in A$ and $\mathbf{1} = (1, \ldots, 1) \in A$, and write $\mathbf{0}_{-i} \in A_{-i}$ and $\mathbf{1}_{-i} \in A_{-i}$ 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

[5] A potential is a function on action profiles and states, with the property that at each state, each player's payoff gain from switching actions is equal to the corresponding gain in the value of this function.

[6] A function on action profiles and states is convex if, at each state, the value of any action profile is smaller than the convex combination of those of all players choosing the low action and all players choosing the high action, with the weight being the fraction of high action players in the action profile in consideration.

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

requires that there exist a state where action 1 is a dominant action for all players: that is, there exists $\overline{\theta} \in \Theta$ 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

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

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 with all players initially choosing action 0.[10] 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 $\theta \in \Theta$.

1.2. *Implementability*

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: *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.

[8]This assumption will be maintained throughout the analysis and used in Theorem A.1(2) (and results that use Theorem A.1(2)). This form of the assumption, however, is stronger than needed. See Appendix B.2.6 of the Supplemental Material for a relaxation.

[9]The countability restriction is made for expositional simplicity only. In particular, Theorem A.1(1) holds with possibly uncountable measurable spaces of types; see Appendix B.2.5 of the Supplemental Material.

[10]Because of the countability of type spaces where there is no issue on measurability, the strategy sets are naturally complete lattices and the payoffs are continuous in strategies (in the pointwise convergence topology), so that standard results on supermodular games apply to our setting.

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

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

for all $i \in I$ and $a_i, a_i' \in A_i$. Bergemann and Morris (2016) showed the following:

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 the following:

DEFINITION 2: Outcome ν is *fully implementable* if there exists an information structure $\mathcal{T}$ such that $(\mathcal{T}, \sigma)$ induces ν for all $\sigma \in E(\mathcal{T})$.[11]

And the intermediate notion we study is the following:

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, $FI \subset SI \subset BCE$. Our first main result, Theorem 1 in Section 3, characterizes the closure $\overline{SI}$, while the characterization of SI is given in Appendix A.1. A characterization of FI is reported in Appendix B.1 of the Supplemental Material (Morris, Oyama, and Takahashi (2024)).

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) discussed contracting applications where the smallest equilibrium is the Pareto-efficient equilibrium for the players. Cooper (1994) argued 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.

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

1.3. *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

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

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 $\sum_{a, \theta} \nu(a, \theta) V(a, \theta)$. Our second main result, Theorem 2 in Section 4, will identify an optimal outcome under additional assumptions.

2. A LEADING EXAMPLE

We will use the following example to illustrate ideas throughout the paper. Let us label the action 1 "invest" and the action 0 "not invest." The payoff to not invest is always 0. There are two players, $I = \{1, 2\}$. Player 1 has a cost 7 of investing while player 2 has a cost 8. Each player receives a return 3 to investing when the other player invests, so the game is supermodular. There are two states, **b** ("bad") and **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 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., $V(a, \theta) = |\{i \in I | a_i = 1\}|$ 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 for any partially implementable perfect coordination outcome, the obedience constraint for player 1 (more willing to invest) does not bind, and the gap 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 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	$\frac{1}{2}$

; (2.2)

indeed, the following outcome will be shown to be S-implementable (and in fact 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	$\frac{1}{2}$

, (2.3)

which converges to the former outcome (2.2) as $\delta \to 0$. The expected number of players who invest is $\frac{3}{2}$ under (2.2). 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.[12] 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 (2.3), 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.[13]

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, Perego, and Taneva (2020) (in a symmetric version of this example) and Li, Song, and Zhao (2023) (in regime-change games) are less surprising. The perfect coordination property holds in this example despite the costs being asymmetric. We will see in Section 4 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 use to characterize (the closure of) the set of S-implementable outcomes. 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 best response provided that only players who received the recommendation earlier than they did switch.

[12] In Appendix A.6, we demonstrate that no outcome close to (2.2) can be S-implementable with public signals.

[13] Mathevet, Perego, and Taneva (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.

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

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

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).[14]

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

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

for all $i \in I$.

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* if there exists an ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ that induces ν and satisfies 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 by showing that it is satisfied by outcome (2.2) in our example in Section 2. 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	$\frac{1}{12}$	$\frac{1}{6}$

.

[14]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 sequential obedience:

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

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

and hence the induced outcome (2.2) also does. Note that, while outcome (2.2) 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, fully characterizes the closure $\overline{SI}$ of the set SI of S-implementable outcomes.

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

In particular, analogously to BCE, $\overline{SI}$ is a convex polytope.[15]

In Appendix A.1, we provide a characterization of SI (Theorem A.1), from which Theorem 1 as well as Corollary 1 below are shown to follow in Appendix A.2.

Theorem 1 requires obedience (necessary for partial implementation) as well as 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 A.1, we can find a first-order stochastically dominating outcome in SI.[16] By a continuity argument, we thus have the following.

COROLLARY 1: *If an outcome ν satisfies consistency and sequential obedience, then there exists an outcome $\hat{\nu}\in\overline{SI}$ that first-order stochastically dominates ν.*

Theorem 1 and Corollary 1 have important implications to the adversarial information design problem (1.2). By Theorem 1, we immediately have the following.

[15]The set of ordered outcomes that satisfy sequential obedience is characterized by a finite system of weak linear inequalities and thus is a convex polytope: by Theorem 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 $\nu(a,\theta)=\sum_{\gamma:a(\gamma)=a}\nu_\Gamma(\gamma,\theta)$.

[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 $a'\geq a$).

COROLLARY 2: *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 sequential obedience.*

By Corollary 1, 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 sequential obedience (without obedience imposed).

In the remainder of this subsection, we sketch a proof of Theorem 1. First consider necessity (i.e., the "only if" part of Theorem 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 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 (with strict inequality): 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. A continuity argument establishes that sequential obedience is satisfied by any outcome in $\overline{SI}$.

Second, consider sufficiency (i.e., the "if" part of Theorem 1). The proof is by construction. Here we illustrate the construction by showing how to S-implement outcome (2.3) 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. Outcome (2.3) is induced by the ordered outcome

	b	**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)

which satisfies sequential obedience with strict inequalities. 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

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

see Table I. 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 $\eta \approx 0$,

TABLE I

INFORMATION STRUCTURE IMPLEMENTING OUTCOME (2.3).

b

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

approximately the payoffs to investing for players 1 and 2 of type $t_i = \tau$ are then greater than (positive multiplications of)

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

and

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

respectively, which are strictly positive by the conditions (3.3) and (3.4). Therefore, by induction, both players of types $t_i < \infty$ invest in the smallest equilibrium. Note that players of type $t_i = \infty$ know that the state is **b** and hence do not invest. Thus, the outcome (2.3) 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)) and on higher order beliefs, in particular on robustness to incomplete information (Rubinstein (1989), Carlsson and van Damme (1993), Kajii and Morris (1997)).

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 action 0), that is, there exists $\gamma \in \Gamma$ such that $a(\gamma) = \underline{a}$ and

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

for all $i \in I$ such that $\underline{a}_i = 1$. In particular, $\mathbf{1}$ (all playing action 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 considered an exogenous initial supermodular game and added transfers in some form (thus determining the payoff functions f_i endogenously) to implement a target outcome as a unique equilibrium. They then asked 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 (possibly contingent on the state θ) and the condition is written as the expectation with respect to the random ordering.[17] Formally, an ordered outcome $\rho \in \Delta(\Gamma)$ satisfies sequential obedience in a complete information game $(f_i)_{i\in I}$ if

$$\sum_{\gamma\in\Gamma_i} \rho(\gamma) f_i\big(a_{-i}(\gamma)\big) \geq 0 \tag{3.6}$$

for all $i \in I$. An outcome $\xi \in \Delta(A)$ satisfies 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 in $(f_i)_{i\in I}$.

By Theorem A.1, 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 $k \to \infty$.[18] Under the maintained assumption of dominance state, we have the following by Theorem A.1 along with an argument similar to the proof of Theorem 1.

COROLLARY 3: *An outcome is limit S-implementable at θ^* if and only if it satisfies obedience and sequential obedience in $(d_i(\cdot, \theta^*))_{i\in I}$.*

[17]This condition has appeared in Oyama and Takahashi (2020) and also played an important role in Halac, Lipnowski, and Rappoport (2021). See also Gershkov and Szentes (2009) for an earlier study where a similar condition is found (but in a different, voting situation).

[18]The latter condition can also be equivalently written as $\nu^k(\cdot, \theta^*) \to \xi$ as $k \to \infty$.

The proof is given in Appendix A.2.

As an illustration, consider the case of two players, and suppose that in the complete information game at θ^*, both **1** and **0** are strict equilibria, so that (the degenerate outcome on) **0** is the S-implementable, but not fully implementable, outcome when $\mu(\theta^*) = 1$. It can be verified that (the degenerate outcome on) **1** satisfies sequential obedience if and only if it is (weakly) risk dominant. Therefore, the "if" part of Corollary 3 implies that if **1** is a risk-dominant equilibrium, then it 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)). Conversely, if **0** is a risk-dominant equilibrium (and hence no other outcome satisfies sequential obedience), then the "only if" part of Corollary 3 implies that as $\mu^k(\theta^*) \to 1$, any information structures induce equilibrium outcomes ν^k such that $\sum_{\theta \in \Theta} \nu^k(\mathbf{0}, \theta) \to 1$, that is, **0** is robust to incomplete information (Kajii and Morris (1997)). Thus, our S-implementability question can be viewed as an incomplete information generalization of the robustness question.[19]

4. APPLICATION TO INFORMATION DESIGN WITH ADVERSARIAL EQUILIBRIUM SELECTION

In this section, we study the optimal information design problem with adversarial equilibrium selection. Under the monotonicity of the designer objective V, it amounts to maximization of (the expectation of) V on the domain $\overline{SI}$ (Section 1.3) and is expressed as a finite-dimensional linear problem (Corollary 2 in Section 3.2). Here, we impose additional restrictions on the base game, which are satisfied in many games found in applications. In Appendix A.4.1, we discuss two classes of such examples, investment games and regime change games.

We assume that the base game is a *potential game*. 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 6: 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 $\theta \in \Theta$,

$$d_i(a_{-i}, \theta) = \Phi\big((1, a_{-i}), \theta\big) - \Phi\big((0, a_{-i}), \theta\big)$$

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 $\theta \in \Theta$. For example, the game (2.1) in Section 2 is a potential game with a potential

b	Not	Invest
Not	0	−8
Invest	−7	−12

g	Not	Invest
Not	0	1
Invest	2	6

. (4.1)

[19] In Morris, Oyama, and Takahashi (2023), we formally described a tight connection between limit implementation by information design and a modified version of the "robustness to incomplete information" notion of Kajii and Morris (1997).

We will now see that the sequential obedience condition can be simplified to a condition with a single inequality 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.

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 $\Phi(\mathbf{1}, \theta)$.[20] Write $n(a)$ for the number of players choosing action 1 in action profile $a \in A$.

DEFINITION 7: Potential Φ satisfies *convexity* if

$$\Phi(a, \theta) \leq \frac{n(a)}{|I|}\Phi(\mathbf{1}, \theta) \tag{4.2}$$

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

For the game (2.1) in Section 2, the potential as given in (4.1) satisfies convexity.

The convexity 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 (4.2) 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

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

Here, $\Delta(a, \theta)$ measures how much higher the value of the potential is for a relative to the average of action 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

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

and so

$$\Delta(a, \theta) \leq M\big(n(a), \theta\big)$$

[20] 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, Morris, and Pauzner (2003), Leister, Zenou, and Zhou (2022)).

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

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

DEFINITION 8: Outcome ν satisfies *perfect coordination* if, for all $\theta \in \Theta$, $\nu(a, \theta) > 0$ only for $a \in \{\mathbf{0}, \mathbf{1}\}$.

This property has been introduced in the context of regime change games by Inostroza and Pavan (2022).

Now we have the following:[21]

PROPOSITION 2: *In a convex potential game, a perfectly coordinated outcome ν satisfies sequential obedience if and only if the average potential of* $\mathbf{1}$ *under ν is nonnegative*:

$$\sum_{\theta \in \Theta} \nu(\mathbf{1}, \theta)\Phi(\mathbf{1}, \theta) \geq 0. \tag{4.3}$$

The proof is given in Appendix A.4.

We now consider the optimal information design problem. In the following, we normalize the designer's objective V so that $V(\mathbf{0}, \theta) = 0$ for all $\theta \in \Theta$. Our main characterization of optimal outcomes requires one additional assumption on V:

DEFINITION 9: 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 $V(a, \theta) = (\frac{n(a)}{|I|})^{\alpha}$ 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 the regime change games; see Example A.2 in Appendix A.4.1.

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 will be shown in Theorem 2, there will be an optimal outcome that satisfies perfect coordination. Once we know that the

[21]For the complete information case with a potential, Segal (2003) showed (in our language) that if the potential of $\mathbf{1}$ in the game obtained by adding transfers is positive, then the deterministic divide-and-conquer condition (3.5) is satisfied for any permutation γ of all players with an appropriate choice of transfers with the same total, and vice versa. The resulting potential satisfies convexity.

solution satisfies perfect coordination, due to Proposition 2 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 sequential obedience (in the form of condition (4.3) in Proposition 2):

$$\max_{(\nu(\mathbf{1},\theta))_{\theta\in\Theta}} \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)V(\mathbf{1},\theta) \tag{4.4a}$$

subject to

$$0 \le \nu(\mathbf{1},\theta) \le \mu(\theta) \quad (\theta \in \Theta), \tag{4.4b}$$

$$\sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)\Phi(\mathbf{1},\theta) \ge 0. \tag{4.4c}$$

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 $\Phi(\mathbf{1},\theta) \ge 0$ 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 $V(\mathbf{1},\theta)$.

Concretely, assume for simplicity that $V(\mathbf{1},\theta) > 0$ for all $\theta \in \Theta$ such that $\Phi(\mathbf{1},\theta) < 0$ (i.e., remove all "bad states" that are irrelevant for the designer), and 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 (with a convention $\frac{x}{0} = \infty$ for $x \ge 0$):[22]

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

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

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

for $\theta \in \{0,1,\dots,|\Theta|\}$. If $\Psi(0) = \sum_{\theta'\in\Theta} \mu(\theta')\Phi(\mathbf{1},\theta') \ge 0$, 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 $\Psi(\theta) \ge 0$ if and only if $\theta \ge \theta^*$. Note that $\Phi(\mathbf{1},\theta^*) < 0$. Let

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

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

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

Now define the perfectly coordinated outcome ν^* by

$$\nu^*(a,\theta)=\begin{cases}\mu(\theta) & \text{if } a=\mathbf{1} \text{ and } \theta>\theta^*,\\ p^* & \text{if } a=\mathbf{1} \text{ and } \theta=\theta^*,\\ \mu(\theta)-p^* & \text{if } a=\mathbf{0} \text{ and } \theta=\theta^*,\\ \mu(\theta) & \text{if } a=\mathbf{0} \text{ and } \theta<\theta^*,\\ 0 & \text{otherwise,}\end{cases} \tag{4.5}$$

which clearly satisfies consistency (4.4b). This outcome satisfies the sequential obedience constraint (4.4c) with equality:

$$\sum_{\theta\in\Theta}\nu^*(\mathbf{1},\theta)\Phi(\mathbf{1},\theta)=\Psi\big(\theta^*\big)+p^*\Phi\big(\mathbf{1},\theta^*\big)=0. \tag{4.6}$$

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

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

since by the convexity of Φ, $\Phi((1,\mathbf{0}_{-i}),\theta)\le\frac{1}{|I|}\Phi(\mathbf{1},\theta)<0$ for all $\theta\le\theta^*$. Thus, $\nu^*\in\overline{SI}$ by Proposition 2 and Theorem 1. Theorem 2 shows that ν^* is an optimal solution to the problem (4.4) and that it is an optimal outcome of the adversarial information design problem.

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* (4.5) *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, that is, $V(a,\theta)=n(a)$, so that restricted convexity is satisfied. With the potential Φ given in (4.1) (which satisfies convexity), we have $\Psi(0)=\sum_{\theta'\in\{\mathbf{b},\mathbf{g}\}}\mu(\theta')\Phi(\mathbf{1},\theta')=-3$ and $\Psi(\mathbf{b})=\mu(\mathbf{g})\Phi(\mathbf{1},\mathbf{g})=3$ (and $\Psi(\mathbf{g})=0$ by convention), and hence $\theta^*=\mathbf{b}$ is the threshold state. With $p^*=\frac{1}{4}$, the optimal outcome ν^* is thus as given in (2.2) in Section 2.

The characterization of the optimal solution as given in Theorem 2 becomes yet simpler in a continuous version of our problem with a continuum of symmetric players and a continuous state space $\Theta\subset\mathbb{R}$. Assume that 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, and also that V satisfies restricted convexity. In this version, 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 (4.5) 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

$$\int_{\theta>\theta^*}\Phi(1,\theta)\,d\mu(\theta)=0 \tag{4.7}$$

(with μ denoting the probability distribution of θ also in this case).[23]

5. RELATED LITERATURE

5.1. *Higher Order Beliefs and Robustness to Incomplete Information*

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) generalized 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. In Section 3.3, we discussed the formal implications of our results in the present paper—in particular for the limit complete information case—to the higher order beliefs literature.

5.2. *Contracting With Externalities*

Our approach has been to fix the base game payoffs $(d_i)_{i\in I}$ and show that sequential obedience characterizes the set of outcomes that are S-implementable. Alternatively, if the outcome to be smallest equilibrium implemented has all players choose the high action, the sequential obedience condition can read as characterizing the set of payoffs for which playing the high action is uniquely rationalizable. This interpretation of our results provides a new perspective on contracting with externalities (Segal (2003) and Winter (2004)), where agents make decisions about whether to participate or not in the presence of strategic complementarities.[24] In fact, the recent studies by Halac, Lipnowski, and Rappoport (2021) and Moriya and Yamashita (2020), who considered the optimal joint design of transfers and information and showed 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, can be understood from this viewpoint. In particular, the optimization problem of Halac, Lipnowski, and Rappoport (2021) can be equivalently rewritten as a problem with the constraint that "all participating" is limit S- (hence fully) implementable, which is characterized by our sequential obedience condition—a stochastic version of "divide-and-conquer." In Morris, Oyama, and Takahashi (2022b), we formally describe the above solution method. By appealing also to the fact that their model has a potential, we obtain additional insights over Halac, Lipnowski, and Rappoport (2021).[25]

[23] In Morris, Oyama, and Takahashi (2022a), we gave a formal derivation of this result.

[24] In Segal (2003), agents' action spaces are designed by the principal who must allow non-participation (the lowest action), in which case all actions above the smallest equilibrium can be eliminated, and hence actions may in effect be viewed as binary.

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

5.3. *Information Design With Adversarial Equilibrium Selection*

In this literature, three papers are most relevant.[26] 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, Perego, and Taneva (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, Song, and Zhao (2023) 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, Perego, and Taneva (2020) and Li, Song, and Zhao (2023). However, all these papers assume symmetric payoffs, where the perfect coordination property is less surprising.[27] The perfect coordination property is more surprising in games with asymmetric payoffs. We illustrated this point in Section 2 with an asymmetric version of the example of Mathevet, Perego, and Taneva (2020). There is generally multiplicity in implementing information structures. The information structures implementing the optimal outcome in Mathevet, Perego, and Taneva (2020) and Li, Song, and Zhao (2023) are tailored to the applications, whereas our implementing information structure construction applies to general BAS games. In Morris, Oyama, and Takahashi (2022a), for the continuous limit model as described after Theorem 2 (which encompasses the regime change game studied by Li, Song, and Zhao (2023) as a special case), we provide another, simple global game implementation of the optimal outcome as given by (4.7) which applies to all symmetric and state-monotonic 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, Perego, and Taneva (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, Shin, and Yildiz (2016), respectively). We chose 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 anal-

[26]Earlier, Kamien, Tauman, and Zamir (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) illustrated the implications of the higher order beliefs literature for information design.

[27]Inostroza and Pavan (2022, Additional Material) showed 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.3, we show that the perfect coordination property holds in asymmetric regime change games within our setting as well.

ogy with the partial implementation case, and reduces the original infinite-dimensional problem into a finite-dimensional linear program.

6. 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 Material (Morris, Oyama, and Takahashi (2024)).

6.1. *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.1 that the arguments for full implementation are straightforward extensions of the results for S-implementation. An outcome is fully implementable only 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 dominance state assumption, 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.

6.2. *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.2.1 for formal arguments.

6.3. *Many Actions*

For (supermodular) games with more than two actions, a natural generalization of the sequential obedience 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 an incentive to do so

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

when expecting that the switches before the current switch have occurred; see Supplemental Appendix B.2.2 for a formal account. Then, the necessity of this condition for S-implementability can be proved almost identically as in the proof of Theorem A.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, we do not expect the generalized sequential obedience condition (along with consistency, obedience, and an appropriately modified version of dominance state) to be sufficient for S-implementability in all games. A proof strategy of the same approach as in the proof of Theorem A.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.2.2, 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]

6.4. *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 information sharing.[30] In Supplemental Appendix B.2.3, 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.

6.5. *Finite Information Structures*

Our implementation in the proof of Theorem A.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.2.4 for a formal argument. Specialized to a symmetric two-player two-state example, Mathevet, Perego, and Taneva (2020) presented 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

[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.

[30] Galperti and Perego (2020) studied 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) considered a similar model but accounted for incentives. Both papers consider partial implementation.

open problem to characterize the smallest number of types needed to implement a given outcome for general BAS games.

APPENDIX A

A.1. *Characterization of S-Implementability*

In this section, we provide a necessary and essentially sufficient condition for S-implementability in terms of a strict version of sequential obedience.

DEFINITION A.1: An ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ satisfies *strict sequential obedience* if

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) d_i\big(a_{-i}(\gamma), \theta\big) > 0 \tag{A.1}$$

for all $i \in I$ such that $\nu_\Gamma(\Gamma_i \times \Theta) > 0$.

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

Our sufficiency holds only for outcomes where all players choose action 1 with positive probability at the dominance state $\overline{\theta}$.[31]

DEFINITION A.2: Outcome ν satisfies *grain of dominance* if $\nu(\mathbf{1}, \overline{\theta}) > 0$.

The set of S-implementable outcomes is characterized as follows:

THEOREM A.1:
(1) *If an outcome is S-implementable, then it satisfies consistency, obedience, and strict sequential obedience.*
(2) *If an outcome satisfies consistency, obedience, strict sequential obedience, and grain of dominance, then it is S-implementable.*

In the subsequent subsections, we prove the sufficiency part (part (1)) and the necessity part (part (2)) of Theorem A.1, respectively. There, 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 A.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,

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

[31] See Supplemental Appendix B.2.7 for the indispensability of this condition.

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

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 strict 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

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

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

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

Thus, ν satisfies strict sequential obedience.

A.1.2. *Proof of Theorem A.1(2)*

Let $\nu \in \Delta(A \times \Theta)$ satisfy consistency, obedience, strict sequential obedience, and grain of dominance, and let $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ be an ordered outcome establishing strict 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 strict sequential obedience, that is,

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

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

$$\bar{q}d_i(\mathbf{0}_{-i},\overline{\theta})+(1-\bar{q})\min_{\theta\neq\overline{\theta}}d_i(\mathbf{0}_{-i},\theta)>0 \tag{A.2}$$

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

$$\frac{\dfrac{\varepsilon}{|I|-1}}{\dfrac{\varepsilon}{|I|-1}+\eta}\geq\bar{q} \tag{A.3}$$

and

$$\sum_{\gamma\in\Gamma_i,\theta\in\Theta}(1-\eta)^{|I|-n(a_{-i}(\gamma))-1}\tilde{\nu}_\Gamma(\gamma,\theta)d_i\big(a_{-i}(\gamma),\theta\big)>0 \tag{A.4}$$

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 $\sum_{t_{-i},\theta} \pi(t_i=\tau, t_{-i}, \theta) = \frac{\varepsilon}{|I|-1} + \sum_{\gamma:\ell(i,\gamma)\leq\tau,\theta}(1-\varepsilon)\eta(1-\eta)^{\tau-\ell(i,\gamma)}\tilde{\nu}_\Gamma(\gamma,\theta) \leq \frac{\varepsilon}{|I|-1} + (1-\varepsilon)\eta \sum_{\gamma:\ell(i,\gamma)\leq\tau,\theta}\tilde{\nu}_\Gamma(\gamma,\theta) \leq \frac{\varepsilon}{|I|-1} + \eta$, while the second inequality is by (A.3). *Q.E.D.*

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 $\tau \in \{|I|, |I|+1, \ldots\}$,*

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

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

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

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

as claimed. *Q.E.D.*

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

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

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

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

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

We are in a position to conclude the proof of Theorem A.1(2). 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.2). For $\tau \geq |I|$, suppose that action 1 is uniquely rationalizable for all players of types $t_i \leq \tau - 1$. Then the expected payoff for a player i of type $t_i = \tau$ from playing action 1 is no smaller than

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

where the equality is by Claim A.2 and the inequality by the "perturbed" strict sequential obedience condition (A.4). 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

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

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. *Proofs of Theorem 1 and Corollaries 1 and 3*

A.2.1. *Proof of Theorem 1*

The "only if" part follows from Theorem A.1(1) by a continuity argument. To prove the "if" part, let $\nu \in \Delta(A \times \Theta)$ satisfy consistency, obedience, and 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, strict sequential obedience with, say, $\underline{\nu}_\Gamma \in \Delta(\Gamma \times \Theta)$, and grain of dominance.[32] Then define $\nu^\varepsilon \in \Delta(A \times \Theta)$ by $\nu^\varepsilon = (1-\varepsilon)\nu + \varepsilon\underline{\nu}$. Clearly, ν^ε satisfies consistency, obedience, strict sequential obedience with $(1-\varepsilon)\nu_\Gamma + \varepsilon\underline{\nu}_\Gamma$, and grain of dominance. Hence, we have $\nu^\varepsilon \in SI$ by Theorem A.1(2). Since $\nu^\varepsilon \to \nu$ as $\varepsilon \to 0$, we therefore have $\nu \in \overline{SI}$.

A.2.2. *Proof of Corollary 1*

First, we claim that for any $\nu \in \Delta(A \times \Theta)$ that satisfies consistency, strict 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 A.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.

[32] 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 strict sequential obedience by Theorem A.1(1), and grain of dominance by the assumptions of dominance state and full support.

Now let $\nu \in \Delta(A \times \Theta)$ satisfy consistency and sequential obedience. Then, as in the proof of Theorem 1, there exists a sequence of outcomes $\nu^\varepsilon \in \Delta(A \times \Theta)$ converging to ν that satisfy consistency, strict 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 Theorem 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.

A.2.3. *Proof of Corollary 3*

The "only if" part follows from Theorem A.1(1) by a continuity argument. To prove the "if" part, let $\xi \in \Delta(A)$ satisfy obedience and 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 any ordered outcome such that $\bar{\rho}(\bar{\gamma}) = 1$ for some sequence $\bar{\gamma}$ of all players. By the dominance state assumption, $\bar{\rho}$ satisfies strict sequential obedience in $(d_i(\cdot, \overline{\theta}))_{i \in I}$. 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 $\nu_\Gamma^k(\cdot, \theta^*) = (1 - \frac{1}{k})\rho$ 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, strict sequential obedience, and grain of dominance. Hence, $\nu^k \in SI(\mu^k)$ by Theorem A.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^*.

A.3. *A Dual Representation of Sequential Obedience*

In this section, we 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 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 A.1 below. It will be used to prove Proposition A.2 in Section A.4.

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:

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

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

PROPOSITION A.1: *An outcome ν satisfies sequential obedience* (*resp. strict sequential obedience*) *if and only if, for any $(\lambda_i)_{i \in I} \in \mathbb{R}_+^I$ such that $\lambda_i > 0$ for some $i \in I(\nu)$,*

$$\sum_{a \in A, \theta \in \Theta} \nu(a, \theta) \max_{\gamma : a(\gamma) = a} \sum_{i \in S(a)} \lambda_i d_i\big(a_{-i}(\gamma), \theta\big) \geq (\textit{resp.} >) 0. \tag{A.5}$$

Thus, sequential obedience requires that for any player weights, the expected weighted sum of payoff changes along the best path be nonnegative.

For illustration, consider outcome (2.3) in the example in Section 2. For given $(\lambda_1, \lambda_2) \in \mathbb{R}_+^2 \setminus \{(0, 0)\}$, the left-hand side of (A.5) is computed as

$$\left(\frac{1}{4} - \delta\right) \max\big\{\lambda_1 \times (-7) + \lambda_2 \times (-5), \lambda_2 \times (-8) + \lambda_1 \times (-4)\big\}$$
$$+ \frac{1}{2} \max\{\lambda_1 \times 2 + \lambda_2 \times 4, \lambda_2 \times 1 + \lambda_1 \times 5\}$$

$$= \begin{cases} (\lambda_2 - \lambda_1)\left(\frac{3}{4} + 5\delta\right) + \lambda_1(12\delta) & \text{if } \lambda_1 \le \lambda_2, \\ (\lambda_1 - \lambda_2)\left(\frac{3}{2} + 4\delta\right) + \lambda_2(12\delta) & \text{if } \lambda_1 \ge \lambda_2, \end{cases}$$

which is always nonnegative (resp. positive) if $\delta = 0$ (resp. if $\delta > 0$). Thus, Proposition A.1 guarantees the existence of some ordered outcome that induces outcome (2.3) and satisfies sequential obedience (resp. strict sequential obedience) if $\delta = 0$ (resp. if $\delta > 0$).

PROOF OF PROPOSITION A.1: Given any $\nu \in \Delta(A \times \Theta)$, let $N_\Gamma(\nu) = \{\nu_\Gamma \in \Delta(\Gamma \times \Theta) | \sum_{\gamma: a(\gamma)=a} \nu_\Gamma(\gamma, \theta) = \nu(a, \theta)\}$ and $\Lambda(\nu) = \{\lambda \in \Delta(I) | \sum_{i \in I(\nu)} \lambda_i = 1\}$, which are each convex and compact. For $\nu_\Gamma \in N_\Gamma(\nu)$ and $\lambda \in \Lambda(\nu)$, let

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

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

$$\max_{\nu_\Gamma \in N_\Gamma(\nu)} \min_{\lambda \in \Lambda(\nu)} D(\nu_\Gamma, \lambda) \ge \text{(resp. } >)0. \tag{A.6}$$

Second, (LHS of (A.5)) $= \max_{\nu_\Gamma \in N_\Gamma(\nu)} D(\nu_\Gamma, \lambda)$ for each $\lambda \in \Lambda(\nu)$. Hence, ν satisfies condition (A.5) if and only if

$$\min_{\lambda \in \Lambda(\nu)} \max_{\nu_\Gamma \in N_\Gamma(\nu)} D(\nu_\Gamma, \lambda) \ge \text{(resp. } >)0. \tag{A.7}$$

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.6) holds if and only if (A.7) holds. *Q.E.D.*

A.4. *Sequential Obedience in Potential Games*

In this section, we provide a simpler characterization of sequential obedience for potential games (Proposition A.2), from which Proposition 2 and Theorem 2 in Section 4 are proved under the convexity assumptions. We also discuss two examples of potential games, investment games and regime change games, to illustrate our assumptions.

For any outcome $\nu \in \Delta(A \times \Theta)$, define

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

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 A.2: *In a potential game, an outcome satisfies sequential obedience (resp. strict sequential obedience) if and only if*

$$\Phi_\nu(\mathbf{1}) \geq (\textit{resp.} >)\Phi_\nu(a) \tag{A.8}$$

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

The proof is given in Section A.4.2, where we verify that condition (A.8) is equivalent to the condition given in Proposition A.1. 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 (2.3), 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 (A.8) with weak (resp. strict) 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 (A.8) reduces to the condition that

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

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 3 and Proposition A.2, $\mathbf{1}$ is limit S-implementable at θ^* if and only if it is a weak potential maximizer at θ^*.[33]

[33] In Morris, Oyama, and Takahashi (2022b), we reported interesting connections between the sequential obedience condition in complete information potential games and some well-known concepts from cooperative

A.4.1. *Examples*

Recall that $n(a) = |S(a)|$ denotes the number of players choosing action 1 in action profile $a \in A$, and (abusing notation slightly) we also let $n(a_{-i})$ denote the number of players choosing action 1 in action profile $a_{-i} \in A_{-i}$.

EXAMPLE A.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:

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

It satisfies convexity if and only if

$$\frac{1}{\ell} \sum_{k=1}^{\ell} (h_k - c_k) \leq \frac{1}{|I|} \sum_{k=1}^{|I|} (h_k - c_k) \tag{A.9}$$

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.

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 $c_2 = 8$. Its potential, as given in (4.1) in Section 4, satisfies convexity, where $h_1 - c_1(= -7) < h_2 - c_2(= -5)$.

EXAMPLE A.2—Regime Change Game: Let $\Theta = \{1, \ldots, |\Theta|\}$, and

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

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.

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 $i \in S$).

The regime collapses if the number of attackers (action 0 players) is larger than $k(\theta)$, or equivalently, the number of non-attackers (action 1 players) is smaller than $|I| - k(\theta)$. Given $a_{-i} \in A_{-i}$, attack (action 0) yields a gross benefit 1 (resp. 0) upon regime change, that is, if $n(a_{-i}) < |I| - k(\theta)$ (resp. upon status quo, i.e., if $n(a_{-i}) \geq |I| - k(\theta)$), with cost c_i, while the payoff of abstention (action 1) is always 0. 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, Song, and Zhao (2023).

This game has a potential:

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

It satisfies convexity if and only if $c_1 = \cdots = c_{|I|}$.

Suppose that the designer's objective is to maximize the probability of maintaining the status quo:[34]

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

Since $\Phi(a, \theta) > \Phi(\mathbf{1}, \theta)$ holds only when $n(a) < |I| - k(\theta)$ (i.e., when the regime collapses), this objective function V satisfies restricted convexity.

A.4.2. *Proof of Proposition A.2*

Suppose that the base game admits a potential Φ. By Proposition A.1, it suffices to show that $\nu \in \Delta(A \times \Theta)$ satisfies condition (A.5) in Proposition A.1 if and only if it satisfies condition (A.8) in Proposition A.2.

The "only if" part: Suppose that ν satisfies sequential obedience (resp. strict sequential obedience) and hence condition (A.5). Fix any $a \in A$ such that $S(a) \subsetneqq I(\nu)$. 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 $\gamma' = (i_1, \ldots, i_{|S(a')|})$ be any such sequence, where $\{i_1, \ldots, i_{|S(a') \cap S(a)|}\} = S(a') \cap S(a)$. Thus, we have

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

[34]This objective is studied in the regime change applications of Inostroza and Pavan (2022) and Li, Song, and Zhao (2023) (Inostroza and Pavan (2022) also considered some more general objectives).

Therefore, we have

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

which is nonnegative (resp. positive) by condition (A.5).

The "if" part: Suppose that ν satisfies condition (A.8). We want to show that ν satisfies condition (A.5). Fix any $(\lambda_i)_{i \in I} \in \mathbb{R}_+^I$ such that $\lambda_i > 0$ for some $i \in I(\nu)$. Let $\gamma^\lambda = (i_1, \dots, i_{|I|})$ be a permutation of all players such that $\{i_1, \dots, i_{|I(\nu)|}\} = I(\nu)$ and $\lambda_{i_1} \leq \cdots \leq \lambda_{i_{|I(\nu)|}}$. Then, we have

$$\begin{aligned}&\text{(LHS of (A.5))} \\ &\geq \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta) \sum_{i \in S(a')} \lambda_i\big(\Phi\big((1, a_{-i}(\gamma^\lambda)\big) \wedge a', \theta\big) - \Phi\big((0, a_{-i}(\gamma^\lambda)\big) \wedge a', \theta\big)\big) \\ &= \sum_{i \in I} \lambda_i \sum_{a' \in A, \theta \in \Theta} \nu(a', \theta)\big(\Phi\big((1, a_{-i}(\gamma^\lambda)\big) \wedge a', \theta\big) - \Phi\big((0, a_{-i}(\gamma^\lambda)\big) \wedge a', \theta\big)\big) \\ &= \sum_{i \in I} \lambda_i\big(\Phi_\nu\big(1, a_{-i}(\gamma^\lambda)\big) - \Phi_\nu\big(0, a_{-i}(\gamma^\lambda)\big)\big) \\ &= \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}}) \sum_{\ell=k}^{|I|} \big(\Phi_\nu\big(1, a_{-i_\ell}(\gamma^\lambda)\big) - \Phi_\nu\big(0, a_{-i_\ell}(\gamma^\lambda)\big)\big) \\ &= \sum_{k=1}^{|I|} (\lambda_{i_k} - \lambda_{i_{k-1}})\big(\Phi_\nu(\mathbf{1}) - \Phi_\nu(\mathbf{1}_{\{i_1, \dots, i_{k-1}\}})\big),\end{aligned}$$

which is nonnegative (resp. positive) by condition (A.8) as desired, where we set $\lambda_{i_0} = 0$.

A.4.3. *Proof of Proposition 2*

By Proposition A.2, sequential obedience is equivalent to condition (A.8) (with weak inequality) 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

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

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 (4.4b), sequential obedience (4.4c), 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 (4.4). Let $(\nu(\mathbf{1},\theta))_{\theta\in\Theta}$ be such that $0\le\nu(\mathbf{1},\theta)\le\mu(\theta)$ and $\sum_{\theta\in\Theta}\nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta)<\sum_{\theta\in\Theta}\nu(\mathbf{1},\theta)\times V(\mathbf{1},\theta)$. For simplicity, we assume that $V(\mathbf{1},\theta)>0$ for all $\theta\in\Theta$.[35] 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 $\sum_{\theta'\ge\theta}\xi(\theta')\le\sum_{\theta'\ge\theta}\xi^{**}(\theta')$ 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

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

But we have

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

since the first term in the right-hand side of the equality equals 0 by (4.6), 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 (4.4).

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 (4.4b) and sequential obedience (4.4c) and whose value is no smaller than that of ν. For each (a,θ), define $\alpha(a,\theta)\in[0,1]$ by

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

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

Take any $\nu\in\overline{SI}$. By Theorem 1 and Proposition A.2, ν satisfies consistency and condition (A.8) in Proposition A.2. Define $\nu'\in\Delta(A\times\Theta)$ by

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

which satisfies the perfect coordination property. Since ν is consistent with μ, so is ν'. Since ν satisfies condition (A.8), we also have

$$\sum_{\theta\in\Theta}\nu'(\mathbf{1},\theta)\Phi(\mathbf{1},\theta)=\sum_{a\in A,\theta\in\Theta}\alpha(a,\theta)\nu(a,\theta)\Phi(\mathbf{1},\theta)$$

[35] Otherwise, define $(\nu'(\mathbf{1},\theta))_{\theta\in\Theta}$ by $\nu'(\mathbf{1},\theta)=\nu^*(\mathbf{1},\theta)$ if $V(\mathbf{1},\theta)=0$ and $\nu'(\mathbf{1},\theta)=\nu(\mathbf{1},\theta)$ otherwise. Then the following argument will go through with ν' in place of ν.

$$\geq \sum_{a\in A,\theta\in\Theta} \nu(a,\theta)\Phi(a,\theta) = \Phi_\nu(\mathbf{1}) \geq 0.$$

Therefore, ν' satisfies (4.4c). For the value of the objective function, we have

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

This completes the proof of Theorem 2.

A.6. *S-Implementation With Public Signals*

In this section, we discuss implementation with public signals for the example from Section 2. The game (2.1) is a special case of the investment game, where its potential satisfies convexity (Example A.1). The designer wants to maximize the expected number of players who choose action 1, that is, $V(a,\theta) = n(a)$ for all $a \in A$ and $\theta \in \Theta$, so that restricted convexity is satisfied.

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	$9q-7, 0$	$9q-4, 9q-5$

, (A.10)

which has a convex potential

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

.

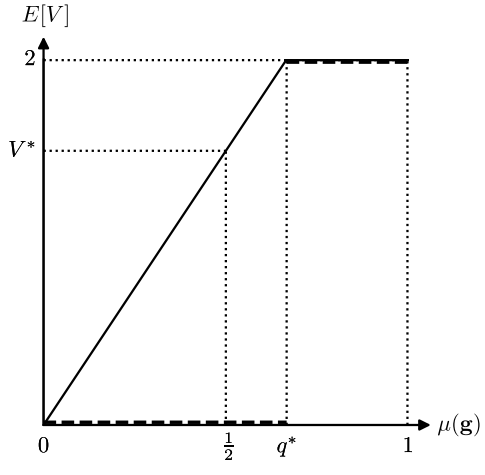

FIGURE A.1.—Optimal values: concavification.

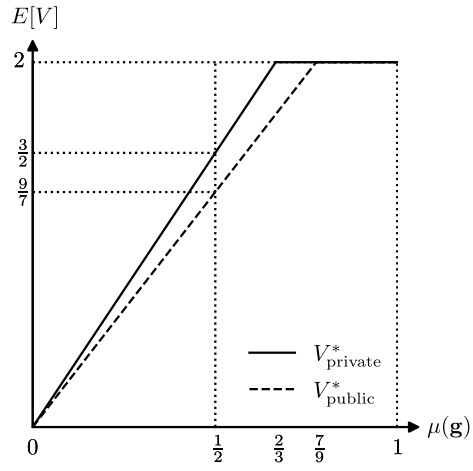

FIGURE A.2.—Optimal values: comparison.

Note that due to the convexity of the potential, pure equilibria of this game are always either (Not Invest, Not Invest) or (Invest, Invest), or both. The profile (Invest, Invest) 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 S-implementation with public signals as a function of $\mu(\mathbf{g})$ is given by the solid line segments in Figure A.1 with $q^* = \frac{7}{9}$ and $V^* = \frac{9}{7}$. When $\mu(\mathbf{g}) = \frac{1}{2}$ $(= \frac{5}{14} \times 0 + \frac{9}{14} \times \frac{7}{9})$ as in Section 2, the optimal outcome under S-implementation with public signals is approached, as $\delta \to 0$, by

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

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 "strict 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 (A.10) if and only if $q \geq \frac{2}{3}$. Indeed, if $\mu(\mathbf{g}) = q \geq \frac{2}{3}$, the ordered outcome given by $\nu_\Gamma(12, \mathbf{b}) = \frac{2}{3}(1 - q)$, $\nu_\Gamma(21, \mathbf{b}) = \frac{1}{3}(1 - q)$, $\nu_\Gamma(12, \mathbf{g}) = \frac{2}{3}q$, and $\nu_\Gamma(21, \mathbf{g}) = \frac{1}{3}q$ (and $\nu_\Gamma(\gamma, \theta) = 0$ otherwise) satisfies 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 sequential obedience and induces the optimal outcome (2.2) 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 A.1). Figure A.2 depicts the optimal values under S-implementation with private signals V^*_{private} (solid line) and S-implementation with public

signals V^*_{public} (dashed line). In particular, when $\mu(\mathbf{g}) = \frac{1}{2}$, no outcome close to outcome (2.2) can be S-implementable with public signals.

REFERENCES

ARIELI, ITAI, AND YAKOV BABICHENKO (2019): "Private Bayesian Persuasion," *Journal of Economic Theory*, 182, 185–217. [780]

AUMANN, ROBERT J. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96. [779]

BERGEMANN, DIRK, AND STEPHEN MORRIS (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures," *Theoretical Economics*, 11, 487–522. [775,776,779]

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95. [795]

CARLSSON, HANS, AND ERIC VAN DAMME (1993): "Global Games and Equilibrium Selection," *Econometrica*, 61, 989–1018. [787,789,794]

CARROLL, GABRIEL (2016): "Informationally Robust Trade and Limits to Contagion," *Journal of Economic Theory*, 166, 334–361. [795]

COOPER, RUSSELL (1994): "Equilibrium Selection in Imperfectly Competitive Economies With Multiple Equilibria," *Economic Journal*, 104, 1106–1122. [779]

FRANKEL, DAVID M., STEPHEN MORRIS, AND ADY PAUZNER (2003): "Equilibrium Selection in Global Games With Strategic Complementarities," *Journal of Economic Theory*, 108, 1–44. [790]

GALPERTI, SIMONE, AND JACOPO PEREGO (2020): "Information Systems," Report. [797]

GERSHKOV, ALEX, AND BALÁZS SZENTES (2009): "Optimal Voting Schemes With Costly Information Acquisition," *Journal of Economic Theory*, 144, 36–68. [788]

GOSSNER, OLIVIER, AND RAFAEL VEIEL (2022): "Rationalizable Outcomes in Games With Incomplete Information," Report. [797]

HALAC, MARINA, ELLIOT LIPNOWSKI, AND DANIEL RAPPOPORT (2021): "Rank Uncertainty in Organizations," *American Economic Review*, 111, 757–786. [788,794,796]

——— (2022): "Addressing Strategic Uncertainty With Incentives and Information," *AEA Papers and Proceedings*, 112, 431–437. [796]

HARSANYI, JOHN C., AND REINHARD SELTEN (1988): *A General Theory of Equilibrium Selection in Games*. Cambridge: MIT Press. [782]

HOSHINO, TETSUYA (2022): "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," *International Economic Review*, 63, 755–776. [795,797]

INOSTROZA, NICOLAS, AND ALESSANDRO PAVAN (2022): "Adversarial Coordination and Public Information Design," Report. [791,795,807]

KAJII, ATSUSHI, AND STEPHEN MORRIS (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65, 1283–1309. [782,787,789,794,795]

KAMENICA, EMIR, AND MATTHEW GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615. [795]

KAMIEN, MORTON I., YAIR TAUMAN, AND SHMUEL ZAMIR (1990): "On the Value of Information in a Strategic Conflict," *Games and Economic Behavior*, 2, 129–153. [795]

LEISTER, C. MATTHEW, YVES ZENOU, AND JUNJIE ZHOU (2022): "Social Connectedness and Local Contagion," *Review of Economic Studies*, 89, 372–410. [790]

LI, FEI, YANGBO SONG, AND MOFEI ZHAO (2023): "Global Manipulation by Local Obfuscation," *Journal of Economic Theory*, 207, 105575. [782,795,807]

MATHEVET, LAURENT, AND INA TANEVA (2022): "Organized Information Transmission," Report. [797]

MATHEVET, LAURENT, JACOPO PEREGO, AND INA TANEVA (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404. [782,795,797]

MERTENS, JEAN-FRANÇOIS, AND SHMUEL ZAMIR (1985): "Formulation of Bayesian Analysis for Games With Incomplete Information," *International Journal of Game Theory*, 14, 1–29. [795]

MILGROM, PAUL, AND JOHN ROBERTS (1990): "Rationalizability, Learning, and Equilibrium in Games With Strategic Complementarities," *Econometrica*, 58, 1255–1277. [776]

MONDERER, DOV, AND DOV SAMET (1989): "Approximating Common Knowledge With Common Beliefs," *Games and Economic Behavior*, 1, 170–190. [795]

MONDERER, DOV, AND LLOYD S. SHAPLEY (1996): "Potential Games," *Games and Economic Behavior*, 14, 124–143. [777,782]

MORIYA, FUMITOSHI, AND TAKURO YAMASHITA (2020): "Asymmetric-Information Allocation to Avoid Coordination Failure," *Journal of Economics & Management Strategy*, 29, 173–186. [794]

MORRIS, STEPHEN, AND HYUN SONG SHIN (1998): "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 88, 587–597. [807]

——— (2004): "Coordination Risk and the Price of Debt," *European Economic Review*, 48, 133–153. [807]

——— (2007): "Common Belief Foundations of Global Games," Report. [795]

MORRIS, STEPHEN, DAISUKE OYAMA, AND SATORU TAKAHASHI (2022a): "Implementation via Information Design Using Global Games," SSRN 4140792. [794,795]

——— (2022b): "On the Joint Design of Information and Transfers," SSRN 4156831. [794,805]

——— (2023): "Strict Robustness to Incomplete Information," *Japanese Economic Review*, 74, 357–376. [789]

——— (2024): "Supplement to 'Implementation via Information Design in Binary-Action Supermodular Games'," *Econometrica Supplemental Material*, 92, https://doi.org/10.3982/ECTA19149. [776,779,796]

MORRIS, STEPHEN, HYUN SONG SHIN, AND MUHAMET YILDIZ (2016): "Common Belief Foundations of Global Games," *Journal of Economic Theory*, 163, 826–848. [795]

OYAMA, DAISUKE, AND SATORU TAKAHASHI (2020): "Generalized Belief Operator and Robustness in Binary-Action Supermodular Games," *Econometrica*, 88, 693–726. [788,794,795]

RUBINSTEIN, ARIEL (1989): "The Electronic Mail Game: Strategic Behavior Under 'Almost Common Knowledge'," *American Economic Review*, 79, 385–391. [782,787,789,794]

SANDMANN, CHRISTOPHER (2020): "Recursive Information Design," Report. [795]

SEGAL, ILYA (2003): "Coordination and Discrimination in Contracting With Externalities: Divide and Conquer?" *Journal of Economic Theory*, 113, 147–181. [779,787,788,791,794]

VIVES, XAVIER (1990): "Nash Equilibrium With Strategic Complementarities," *Journal of Mathematical Economics*, 19, 305–321. [776]

WINTER, EYAL (2004): "Incentives and Discrimination," *American Economic Review*, 94, 764–773. [787,788,794]

Co-editor Asher Wolinsky handled this manuscript.

Manuscript received 12 November, 2020; final version accepted 2 February, 2024; available online 6 February, 2024.