Implementation via Information Design in Binary-Action Supermodular Games*

Stephen Morris† Daisuke Oyama‡ Satoru Takahashi§

April 5, 2020

Abstract

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

As one application of our result, we show that if the game has a convex potential and an information designer wants players to choose the high action, it is optimal choose a perfect coordination outcome, where either all players choose the high action or all player choose

*We are grateful for the comments of seminar participants at the "Current Topics in Microeconomics Theory" conference at the City University of Hong Kong and at the "Theory and Finance Workshop on Coordination and Information Design" at the People Bank of China School of Finance at Tsinghua University in June 2019; at the 2019 Stony Brook International Conference in Game Theory; the Northwestern Workshop on Computer Science and Economics; and at Waseda, Keio, Tsukuba, Yokohama National, Kobe, Osaka, Singapore Management, Harvard/MIT, Kansai, and Seoul National Universities.

†semorris@mit.edu

‡oyama@e.u-tokyo.ac.jp

§ecsst@nus.edu.sg

the low action. The optimal outcome has all playing the highest action on the largest event where that action profile maximizes the ex ante potential.

1 Introduction

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

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

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

Our first main result is a characterization of smallest equilibrium implementability. The char-

[1]Bergemann and Morris (2019) provide an overview of a now large literature building on this observation. What we are calling the information design problem is a many player generalization of the Bayesian persuasion problem described by Kamenica and Gentzkow (2011); see Kamenica (2019) for a survey of this literature. Bergemann and Morris (2013, 2016) characterized the implementable outcomes in the many player case and noted information design applications. Taneva (2019) suggested the terminology "information design". Information design is a special class of mechanism design problems that Myerson (1991) labelled "communication in games", with the twist that the designer is able to deliver information to the players without having to elicit it.

acterization is closely analogous to the obedience characterization of partial implementation. In particular, it corresponds to a finite linear program. The more demanding criterion of smallest equilibrium implementation gives rise to a more demanding *sequential obedience* constraint. Sequential obedience requires that it is possible for the information designer to choose (perhaps randomly conditioning on the state) an ordering of players in which players are advised to play the high action in such a way that they are *strictly* willing to follow the recommendation *even if they only expect players who received the recommendation before them to choose the high action.*

To see why sequential obedience is necessary, suppose that an outcome is smallest equilibrium implementable. Then there exists an information structure where this outcome is played in the smallest equilibrium. Recall that it is a standard property of supermodular games that the smallest equilibrium can be reached by a myopic best response sequence, where we start with all types of all players choosing the low action and sequentially switch players' types to the high action when it is a strict best response to do so (where the order of switches does not matter).[2] The information structure and myopic best response sequence will induce a probability distribution over states and sequences in which players switch. But now a player will have an incentive to switch when he is told to switch even if he only expects those before him to have switched and has no additional information: this is true because we are averaging across scenarios where the player switches in the myopic best response process. But now we have constructed a probability distribution over sequences of recommendations for which sequential obedience holds.

Direct information structures (where players are given action recommendations) will in general not be sufficient to smallest equilibrium implement an outcome. Nonetheless, we show that if an outcome satisfies sequential obedience, and a dominant state property, requiring that the high action is a dominant action in some state, then we can construct a canonical information structure that implements the outcome in the smallest equilibrium. In this mechanism, a sequence is drawn randomly in a way that supports the sequential obedience condition, and in addition a non-negative integer is drawn randomly according to an almost flat exponential distribution. Each player's type is the sum of the integer and his rank in the sequence. Payoffs are re-arranged to ensure that players with very low types have a dominant action to play the high action. One can then argue by induction that if all types of all players up to k play the high action, type $k+1$ of any player has an incentive to choose the high action: by construction this type will be sure

[2]Milgrom and Roberts (1990).

that players with lower ranks than him are playing the high action so this is true by sequential obedience.

Our characterization of full implementation is an easy extension of our smallest equilibrium implementation result, where a reverse sequential obedience condition is required in addition. We focus on smallest equilibrium implementation because it is simpler to explain and most relevant for applications.

We provide two applications of our main result. First, we consider the problem of a designer who always prefers players to choose the high action but expects the worst (i.e., smallest) equilibrium to be played, and has no instrument other than information design. We review below a recent literature which has studied this question. We show that if the game has a convex potential[3] at each state and the designer's objective satisfies a restricted convexity requirement, then the optimal outcome is perfectly coordinated, where either all players choose the high action or all players choose the low action. Convexity of the potential is equivalent to requiring that payoffs in the game are not too asymmetric.

When these conditions are satisfied, the optimal outcome is easy to characterize. All players choose the high action conditional on the highest probability event with the property that the expected potential on that event from all playing the high action exceeds the expected potential from all playing the low action. We show our result is tight, giving general conditions under which the optimal outcome is no longer a coordinated outcome.

For our second application, we consider what happens if the designer can also offer bonuses to players for choosing the high action, and characterize the optimal bonus and critical information structure.

We focus on two applications to illustrate the usefulness of our sequential obedience characterization of smallest equilibrium implementability. However, in the course of these applications, we develop a set of progressively simpler characterizations of sequential obedience (under additional restrictions on the environment) that we believe should prove useful in many other applications.

1.1 Literature

Our work has its roots in a large literature on the role of higher-order beliefs in games. A prominent and early insight in this literature is that a particular Nash equilibrium of a complete

[3]Monderer and Shapley (1996).

information game can be fully implemented via an "infection argument" (Rubinstein (1989) and Carlsson and van Damme (1993)). In particular, consider a two player two action game of complete information with two strict Nash equilibria. In a symmetric game, a Nash equilibrium is said to be risk dominant if each player's action is a best response to a 50/50 conjecture over the actions of the other player. It is possible to construct an incomplete information game where with probability close to 1, payoffs are given by the complete information game, but nonetheless the unique equilibrium of the incomplete information game has the risk dominant action profile of the complete information game played everywhere. Thus we can fully implement the risk dominant outcome by information design, and a small perturbation to payoffs. The argument extends to games with asymmetric payoffs for the appropriate definition of risk dominance. Generically, a 2×2 game with two strict Nash equilibria has exactly one risk dominant equilibrium.

Kajii and Morris (1997) showed that this argument worked only for the risk dominant Nash equilibrium. It is not possible to construct an infection argument for the risk dominated strict Nash equilibrium. Thus the risk dominated equilibrium cannot be fully implemented. A recent paper of Oyama and Takahashi (2020) characterizes when such an infection argument can be constructed for generic BAS games.[4] These papers addressed the ability to fully implement a particular equilibrium of a complete information game by information design (and a small perturbation to payoffs). Our main result characterizes when it is possible to construct an infection argument that fully implements a general outcome in an incomplete information setting. We do not have perturbed payoffs but rather maintain a dominant state property to initiate the infection argument.[5]

[4]These papers, and a large related literature in between them, were phrased differently and did more. Kajii and Morris (1997) said that a Nash equilibrium of a complete information was "robust to incomplete information" if every incomplete information where payoffs are almost given by that complete information game has a Bayes Nash equilibrium where that Nash equilibrium is almost always played. The infection argument then establishes that the risk dominated equilibrium is not robust. Kajii and Morris (1997) described sufficient conditions for robustness in many action many player games, generalizing risk dominance. Ui (2001) and Morris and Ui (2005) generalized those sufficient conditions to arguments using potential and generalized potential functions. Oyama and Takahashi (2020) showed that a sufficient condition identified by Morris and Ui (2005) ("monotone potential maximizer") was also necessary for robustness in generic BAS games. This paper builds on the necessity argument of Oyama and Takahashi (2020).

[5]Morris and Ui (2019) considers a different incomplete information extension of the literature on "robustness of equilibria to incomplete information" described in the previous footnote. They ask: which equilibria of a fixed

Our first application is to the pure information design problem where the information designer can choose the information structure but anticipates that the worst equilibrium will be played ("information design with adversarial equilibrium selection").[6] Inostroza and Pavan (2019) showed that optimal solutions would satisfy the perfect coordination property (PCP) in a regime change game (an example of a BAS game). Mathevet et al. (2020) solved a two player two action example of the problem, showing that the solution also satisfied the PCP property. Contemporaneously with this paper, Li et al. (2019) solved the information design problem in regime change games. Our results provide tight conditions under which PCP holds in general BAS games and provides a general solution which is simple to calculate and interpret. In contrast to these papers, we show that the PCP property continues to hold even when the games' payoffs are asymmetric across players. This contrasts with the partial implementation case where asymmetric payoffs will lead to asymmetric optimal solutions, as emphasized by Arieli and Babichenko (2019). Mathevet et al. (2020) and Li et al. (2019) report different and simpler information structures that implement optimal outcomes tailored to the applications. We report a canonical information structure that works for all BAS games.

Our second application relates to a literature on inducing effort in binary actions games: Segal (1999), Winter (2004) and Halac et al. (2019). We show how to use our methods to generalize a recent contribution of Moriya and Yamashita (2020).

We note that even though our "sequential obedience" is necessary and sufficient for smallest equilibrium implementation, but there is not physical sequentiality in our problem. Physical sequentiality is the focus of Doval and Ely (2019), who characterize when a designer can partially implement an outcome when the actions and payoffs of the players are fixed but the designer can design both the information structure and the extensive form. The version of obedience that

---

incomplete information game have the property that nearby outcomes arise in all nearby incomplete information games? Morris and Ui (2019) propose a definition and sufficient conditions that generalize those in Morris and Ui (2005) in general games, addressing subtle issues concerning the correct definition of robustness.

[6]Early references include Goldstein and Huang (2016), who restrict attention to public information, and Carroll (2016), who considers a trading game that is not supermodular. Carroll (2016) asks what is the lowest possible social welfare across all information structures in a bilateral trading problem when the best equilibrium is played. This corresponds to the information design problem where the information designer wants to minimize social welfare but anticipates that an adversarial equilibrium will be chosen. Bergemann and Morris (2019) and Hoshino (2018) have earlier noted the tight connection between robustness to incomplete information and information design. See also Sandmann (2020) and Ziegler (2020) for recent contributions in this area.

arises in their problem is less demanding than standard static obedience whereas our sequential obedience condition is stronger than the standard static obedience condition. The obedience condition of Doval and Ely (2019) is weaker than ours both because players assume that later players will follow their recommendations, and they assume that later players behavior will respond to their deviations.

The properties of higher order beliefs under the common prior assumption matter to our analysis. However, the implications of the common prior assumption for higher order beliefs is a subtle topic: the best characterization we have of the common prior assumption as a restriction on higher order beliefs is that it is equivalent to no trade (Morris (1994), Samet (1998) and Feinberg (2000)), but this characterization is not very usable; the elegant recent paper of Arieli et al. (2020) highlights how difficult it is to characterize common prior restrictions on higher-order beliefs. While a number of recent papers have emphasized the use of the universal type space in highlighting the role of higher-order beliefs (e.g., Mathevet et al. (2020) and Inostroza and Pavan (2019)), the different tradition in the higher-order beliefs literature has been to use belief operators (Monderer and Samet (1989)) on arbitrary common prior type spaces to track higher-order beliefs. Morris and Shin (2007) and Morris et al. (2016) used generalized belief operators introducing the importance of rank beliefs, i.e., what probability does a player assign to having a higher expectation than other players. Kajii and Morris (1997) and Oyama and Takahashi (2020) prove results by first establishing properties of higher order beliefs that hold on all common prior type spaces (using belief operators and generalized belief operators respectively) and then use those properties to establish their results. Rank beliefs play an important role in sequential obedience. However, in this paper, we do not perform the intermediate step of stating results about higher order beliefs on common prior type spaces. This is a pedagogical choice, as we prefer to focus on the obedience interpretation of our necessary and sufficient conditions. However, it is important to note that common prior properties higher-order beliefs are implicit in our analysis, and could be made explicit with the language of generalized belief operators.

1.2 Setting

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

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

is weakly increasing in $a_{-i} \in A_{-i}$. We denote $\mathbf{0} = (0, \ldots, 0)$ and $\mathbf{1} = (1, \ldots, 1)$, and write $\mathbf{0}_{-i}$ and $\mathbf{1}_{-i}$ for the action profiles of player i's opponents such that all players $j \neq i$ play 0 and 1, respectively.

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 $i \in I$, and $\pi \in \Delta(T \times \Theta)$ is a common prior over $T \times \Theta$, where we write $T = \prod_i T_i$. A type space is *consistent* if $\sum_t \pi(t, \theta) = \mu(\theta)$ for each $\theta \in \Theta$. We maintain a *dominance state assumption* that says that there exists a state where action 1 is a dominant action for all agents: there exists $\overline{\theta} \in \Theta$ such that $d_i(\mathbf{1}_{-i}, \overline{\theta}) > 0$ for all $i \in I$.

Together with the payoff functions $(u_i)_{i \in I}$, a type space $\mathcal{T}$ 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 $\sigma_i \colon T_i \to \Delta(\{0, 1\})$. A strategy profile $\sigma = (\sigma_i)_{i \in I}$ is a (Bayes Nash) equilibrium of the game $\mathcal{T}$ if

$$\begin{aligned} &\sum_{a,t,\theta} \pi(t, \theta) \left(\prod_j \sigma_j(t_j)(a_j) \right) u_i(a, \theta) \\ &\geq \sum_{a,t,\theta} \pi(t, \theta) (\sigma'_i(t_i)(a_i)) \left(\prod_{j \neq i} \sigma_j(t_j)(a_j) \right) u_i(a, \theta) \end{aligned}$$

for all i and σ'_i. We are interested in induced outcomes, where an outcome is a distribution in $\Delta(A \times \Theta)$. A type space $\mathcal{T}$ and strategy profile σ induce outcome $\nu \in \Delta(A \times \Theta)$:

$$\nu(a, \theta) = \sum_t \pi(t, \theta) \prod_i \sigma_i(t_i)(a_i).$$

We write $BNE(\mathcal{T})$ for the set of Bayes Nash equilibria of the game $\mathcal{T}$. Because the game is supermodular, there will always exist a smallest equilibrium, which is in pure strategies (Milgrom and Roberts (1990)). We write $\underline{\sigma}(\mathcal{T})$ for that smallest pure strategy equilibrium.

1.3 Implementation

We will be interested in which outcomes can be implemented by a suitable choice of information structure. The answer will depend on the equilibrium selection rule. We will focus on three different assumptions about equilibrium selection which will give rise to three different notions of implementation.

Definition 1. Outcome ν is *partially implementable* if there exist an information structure $\mathcal{T}$ and a Bayes Nash equilibrium σ of $\mathcal{T}$ such that $(\mathcal{T}, \sigma)$ induces ν.

We write $BCE \subset \Delta(A \times \Theta)$ for the set of partially implementable outcomes. This implementation problem is well understood. An outcome ν satisfies *obedience* if

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

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

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

Bergemann and Morris (2016) called such outcomes Bayes correlated equilibria since they correspond to one natural generalization of correlated equilibrium of Aumann (1974, 1987) to incomplete information games. Note that BCE is characterized by a finite system of linear inequalities and is thus a convex polytope.

Definition 2. Outcome ν is *smallest equilibrium implementable* if there exists an information structure $\mathcal{T}$ such that $(\mathcal{T}, \underline{\sigma}(\mathcal{T}))$ induces ν.

We write $SI \subset \Delta(A \times \Theta)$ for the set of smallest equilibrium implementable outcomes. We characterize SI and its closure $\overline{SI}$ in Section 3 and apply characterizations in Section 4.

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

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

2 Two State Examples

We will illustrate our definitions and preview our results with a series of examples involving two states. There are two equally likely states, good bad (**b**) and good (**g**). Action 0 is labelled "not invest" and action 1 is labelled "invest". We will normalize the payoff to not invest to 0.

2.1 One Player

First suppose that there is a single player and that the payoff to investing is 2 and -4 in the good and bad state respectively:

Example 1: payoffs

b		**g**	
Invest	-4	Invest	2
Not Invest	0	Not Invest	0

An example of a partially implementable outcome is given in the following table.

b		**g**	
Invest	$\frac{1}{4}$	Invest	$\frac{1}{2}$
Not Invest	$\frac{1}{4}$	Not Invest	0

(1)

In the good state, which occurs with probability $\frac{1}{2}$, the player always invests. In the bad state, the player invests with probability $\frac{1}{2}$. This outcome is obedient because when a player receives a recommendation to invest, she assigns probability $\frac{2}{3}$ to the state being good and is thus indifferent between investing and not investing. So outcome (1)is the best outcome for a designer who wants to maximize the probability of investment, and the example corresponds to the leading example of Kamenica and Gentzkow (2011).

Because this outcome leaves the player indifferent between invest and not invest when told to invest, so there are "multiple equilibria" of the one player game. But the following outcome is fully implementable (and smallest equilibrium implementable) for any $\varepsilon > 0$.

b		**g**	
Invest	$\frac{1}{4} - \varepsilon$	Invest	$\frac{1}{2}$
Not Invest	$\frac{1}{4} + \varepsilon$	Not Invest	0

(2)

Thus although (1) is not fully implementable, it is in the closure of the set of fully implementable outcomes.

2.2 Two Players with Symmetric Payoffs

Now let us consider the simplest extension of the one player example to two players. Let the payoff to investing remain 2 and -4 in the good and bad state respectively, *if the other player* *invests*. But if the other player fails to invest, we subtract 1 from a player's payoffs. Thus payoffs in the game are

incomplete information game payoffs

b	Invest	Not	**g**	Invest	Not
Invest	$-4,-4$	$-5,0$	Invest	$2,2$	$1,0$
Not Invest	$0,-5$	$0,0$	Not Invest	$0,1$	$0,0$

But essentially the same outcome as in the one player case will continue to be partially implementable:

b	Invest	Not	**g**	Invest	Not
Invest	$\frac{1}{4}$	0	Invest	$\frac{1}{2}$	0
Not Invest	0	$\frac{1}{4}$	Not Invest	0	0

If each player anticipates that the other player will invest when he would have invested in the one player case, he will have an incentive to behave as before. This is in fact the best partially implementable outcome for a designer who prefers investment.

However, if we make direct recommendations to the players under this outcome, there is a *strict* equilibrium where both players never invest. By construction, players are indifferent about investing when they thought the other player would do so, so they have a strict incentive to not invest when they think the other player will not invest. Unlike in the single player case, no outcome close to this outcome is fully implementable.

However, Mathevet et al. (2020) have shown that the following outcome is in the closure of the fully implementable outcomes:

b	Invest	Not	**g**	Invest	Not
Invest	$\frac{1}{6}$	0	Invest	$\frac{1}{2}$	0
Not	0	$\frac{1}{3}$	Not	0	0

and it is not possible to fully implement any more investment by information design. This example illustrates the perfect coordination property: it is optimal to have either both players invest or both do not.

Our explanation of this result comes from thinking about the complete information game conditional on both players being told to invest. The complete information game payoffs are

complete information payoffs

	Invest	Not
Invest	$\frac{1}{2}, \frac{1}{2}$	$-\frac{1}{2}, 0$
Not	$0, -\frac{1}{2}$	$0, 0$

This is a game where "Invest, Invest" is (just) risk dominant, so we can fully implement it following the logic of Rubinstein (1989) and Carlsson and van Damme (1993). Kajii and Morris (1997) showed that one cannot fully implement a higher probability of investment in this way.

2.3 Two Players with Asymmetric Payoffs

But the conclusion that the perfect coordination property holds is hardly surprising in a game with symmetric payoffs: what force would lead players not to coordinate in that case? A more interesting case is what happens with asymmetric payoffs. So suppose now that player 2 incurs an extra cost $0 \leq c \leq 2$ whenever he invests:

incomplete information game payoffs

b	Invest	Not
Invest	$-4, -4-c$	$-5, 0$
Not	$0, -5-c$	$0, 0$

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

Thus player 1 is more willing to invest than player 2. Thus it seems intuitive that an information designer wanting to induce the most investment would have player 1 invest more likely than player 2. This turns out to be true in the case of partial implementation. The following outcome is partially implementable and maximizes the sum of the probability that player 1 invests and the probability that player 2 invests, among all partially implementable outcomes:

b	Invest	Not
Invest	$\frac{10-5c}{10(4+c)}$	$\frac{6c}{10(4+c)}$
Not	0	$\frac{10+4c}{10(4+c)}$

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

To see why, observe that the probability that both players invest in the bad state is as high as it can be consistent with the obedience constraint of player 2 being satisfied. But since player 1 is more willing to invest, he can be induced more, i.e., with positive probability when the state is bad and player 2 is not investing.

However, if $c \leq 1$, it turns out that the fully implementable outcome with the most investment is

b	Invest	Not
Invest	$\frac{3-c}{2(9+c)}$	0
Not	0	$\frac{6+2c}{2(9+c)}$

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

and this satisfies the perfect coordination property. Intuitively, because heterogeneity of payoffs is not too large it remains optimal to induce players to invest together despite their asymmetries. And why is this the best fully implementable outcome? Consider the complete information game being player conditional on both being told to invest:

	Invest	Not
Invest	$\frac{1}{2}(1+c), \frac{1}{2}(1-c)$	$-\frac{1}{2}(1-c), 0$
Not	$0, -\frac{1}{2}(1+c)$	$0, 0$

Again, we see that "Invest, Invest" is (just) risk dominant in the ex ante game. Our convexity condition reduces to $c \leq 1$ and this optimal outcome illustrates our general characterization.

3 Smallest Equilibrium Implementation

We first report a strengthening of obedience—which we will call "sequential obedience"—that will be necessary and essentially sufficient for smallest equilibrium implementation. Suppose that players' default action was to play 0 but the information designer recommended a subset of agents to play 1. The designer gave those recommendations sequentially, according to some commonly known distribution on states and sequences of recommendation. When players are advised to play action 1, they will accept the recommendation only if it is a strict best response if provided that only agents who received the recommendation earlier than them switch.

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

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

For each $\gamma \in \Gamma$, we denote by $\bar{a}(\gamma) \in A$ the action profile such that player i plays action 1 if and only if i is listed in γ. We will 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:

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

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).

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

$$\sum_{\gamma \in \Gamma_i, \theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > 0 \tag{3}$$

for all $i \in I$ such that $\nu_\Gamma(\Gamma_i \times \Theta) > 0$. It satisfies *weak sequential obedience* if the strict inequality in (3) is replaced with a weak inequality.

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

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

If action profile $\mathbf{0}$ is a Nash equilibrium at every state, then the smallest equilibrium will have all players choosing action 0. Our sufficiency argument will work only for outcomes where all players choose action 1 with positive probability at a dominant state.

Definition 6. An outcome $\nu \in \Delta(A \times \Theta)$ satisfies *dominance* if $\nu(\mathbf{1}, \overline{\theta}) > 0$.

Theorem 1.

1. *If an outcome is smallest equilibrium implementable, then it satisfies consistency, obedience and sequential obedience.*

2. *If an outcome satisfies consistency, obedience, sequential obedience and dominance, then it is smallest equilibrium implementable.*

The proof of Theorem 1 as well as those of the following corollaries are given in Appendix A.1.

Corollary 1. $\nu \in \overline{SI}$ *if and only if* ν *satisfies consistency, obedience, and weak sequential obedience.*

Corollary 2. *If an outcome* ν *satisfies consistency and weak sequential obedience, then there is an outcome* ν' *that first-order stochastically dominates* ν *and satisfies consistency, obedience and weak sequential obedience.*

In particular, $\overline{SI}$ is a convex polytope.

We conclude this section with a dual representation of the sequential obedience condition. There are a number of independent reasons for reporting it. First, we will appeal to it in the proof of Theorem 3 in Section 3. Second, it provides some alternative intuition for the sequential obedience condition. Third, it may be important in future work. And finally, it highlights the important connection with Oyama and Takahashi (2020). They showed that an equilibrium of a BAS game was robust if and only if it was a monotone potential maximizer. The existence of a monotone potential is the dual of the failure of a version of sequential obedience for the complete information case.

For $\nu \in \Delta(A \times \Theta)$, write $S(\nu) = \{i \in I \mid \nu((1, a_{-i}), \theta) > 0 \text{ for some } a_{-i} \in A_{-i} \text{ and } \theta \in \Theta\}$ and $A(\nu) = \{a \in A \mid S(a) \subset S(\nu)\}$, where $\nu(A(\nu) \times \Theta) = 1$.

Proposition 2. $\nu \in \Delta(A \times \Theta)$ *satisfies sequential obedience (resp. weak sequential obedience) if and only if for any* $(\lambda_i)_{i \in S(\nu)} \geq 0$, $(\lambda_i)_{i \in S(\nu)} \neq 0$,

$$\sum_{a \in A(\nu), \theta \in \Theta} \nu(a, \theta) \max_{\gamma \in \Gamma(a)} \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) > (\textit{resp.} \geq)\ 0. \tag{4}$$

The proof is given in Appendix A.2.

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

4 Applications

4.1 Potential Games

While our characterization of smallest equilibrium implementability applies to all BAS games, many relevant applications work with potential games and we can give easier simpler characterizations of sequential obedience in this case.

Definition 7. The game is a *potential game* if there exists $\Phi : A \times \Theta \to \mathbb{R}$ such that for each $\theta \in \Theta$,

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

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

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

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

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

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

where h_k is increasing in k and $R(\theta)$ is strictly increasing in θ. Assume that $R(|\Theta|) + h_1 > c_i$ for all $i \in I$, so that the dominance state assumption holds with $\bar{\theta} = |\Theta|$. We interpret $d_i(a_{-i}, \theta)$ to be the return to investment (action 1), which is (i) increasing in the state; and (ii) increasing in the proportion of others investing (making the game supermodular). But there are heterogeneous costs of investment; without loss we assume that

$$c_1 \le c_2 \le \cdots \le c_{|I|}.$$

This game has a potential:

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

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

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

where $k\colon \Theta \to \mathbb{N}$ is strictly increasing. We assume that $k(1) \ge 1$ and $k(|\Theta|) = |I|$, so that the dominance assumption holds with $\bar{\theta} = |\Theta|$. The interpretation is that action 0 is to attack the regime while action 1 is to abstain from attacking. The regime collapses if the number of attackers (action 0 players) is greater than or equal to $k(\theta)$, or equivalently, the number of non-attackers (action 1 players) is smaller than $|I| - k(\theta)$. This game has a potential:

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

4.2 Simplifying Sequential Obedience

The purpose of this section is to provide simpler characterizations of sequential obedience. Suppose that a subset of players were able to coordinate a deviation where they would always choose action 0, even if action 1 was recommended. We say that an outcome is *coalitionally obedient* if no subset of players would want to deviate in this way. We say that an outcome is *grand coalitionally obedient* if the set of all players would not want to deviate in this way. It is intuitive that such conditions might be relevant for full implementation, since we are trying to provent the possibility of joint deviations to a bad equilibrium. These conditions are easier to check than sequential obedience.

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

4.2.1 Coalitional Obedience

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

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

where $b = a \wedge a'$ denotes the action profile such that $b_i = 1$ if and only if $a_i = a'_i = 1$. Imagine that players are told to play action 0 or 1 according to ν. The function Φ_ν can be interpreted as a complete information game with a common payoff function Φ_ν given as follows: Players choose actions a, before they receive recommendations a' according to ν. Players who choose action 1 (i.e., those for whom $a_i = 1$) actually play 1 only when recommended so (i.e., $a'_i = 1$), while players who are recommended to play action 0 (i.e., those for whom $a'_i = 0$) are forced to play

0 as "passive players". Thus, the resulting action profile is $a \wedge a'$, yielding a payoff $\Phi(a \wedge a')$ to every player. Taking the expectation with respect to ν leads to $\Phi_\nu(a)$.

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

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

Thus, players in $I \setminus S(\nu)$ are treated as "passive players" with probability one in computing Φ_ν. Let $A(\nu) \subset A$ denote the set of action profiles in which these "passive players" play action 0:

$$A(\nu) = \{a \in A \mid S(a) \subset S(\nu)\},$$

where $S(a) = \{i \in I \mid a_i = 1\}$. Equivalently,

$$A(\nu) = \{a \in A \mid a_i = 0 \text{ for all } i \notin S(\nu)\}.$$

By definition,

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

and $\Phi_\nu(\mathbf{1}) = \Phi_\nu(a)$ for all $a \in A$ such that $a_i = 1$ for all $i \in S(\nu)$, in particular, for $a = \mathbf{1}_{S(\nu)}$. Our coalitional obedience condition will require that players strictly prefer to play 1, or $\mathbf{1}_{S(\nu)}$, to jointly deviating to any $a \in A(\nu) \setminus \{\mathbf{1}_{S(\nu)}\}$.

Definition 8. Outcome ν satisfies *coalitional obedience* (resp. *weak coalitional obedience*) if

$$\Phi_\nu(\mathbf{1}) > (\text{resp.} \geq) \; \Phi_\nu(a) \tag{5}$$

for all $a \in A(\nu) \setminus \{\mathbf{1}_{S(\nu)}\}$.

Our first simplification of sequential obedience is:

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

The intuition is that if the game is a potential game, we can add up the gains from deviating across players.

The proof is given in Appendix A.3.

4.2.2 Grand Coalitional Obedience and Perfect Coordination

Grand coalitional obedience corresponds to the even simpler condition that the coalitional obedience condition holds for the set of all players.

Definition 9. Outcome ν satisfies *grand coalitional obedience* (resp. *weak grand coalitional obedience*) if

$$\Phi_\nu(\mathbf{1}) > (\text{resp. } \geq)\ \Phi_\nu(\mathbf{0}) = 0, \tag{6}$$

or equivalently,

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

Sequential obedience will reduce to grand coalitional obedience when incentives to coordinate are high (Proposition 4).

We say that an outcome is a *perfect coordination outcome* if all players are always choosing the same action.

Definition 10. Outcome ν satisfies perfect coordination if $\nu(a,\theta) > 0 \Rightarrow a \in \{\mathbf{0}, \mathbf{1}\}$.

As we noted in the introduction, this property arises in the work of Inostroza and Pavan (2019) and Li et al. (2019) and will play an important role in our applications.

When should we expect perfection coordination to arise? If players' actions maximized the potential function at each state, the necessary and sufficient condition would be:

$$\arg\max_{a \in A} \Phi(a,\theta) \subset \{\mathbf{0}, \mathbf{1}\}$$

for all $\theta \in \Theta$.[7]

A key property in our analysis will be a cardinal strengthening of this property which requires that $\Phi(a,\theta)$ is always more than a convex combination of $\Phi(\mathbf{0},\theta) = 0$ and $\Phi(\mathbf{1},\theta)$.

Definition 11. Potential Φ satisfies *convexity* if

$$\Phi(a,\theta) \leq \frac{n(a)}{|I|}\Phi(\mathbf{1},\theta) \tag{7}$$

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

[7]Frankel et al. (2003) show that this is a sufficient condition for a perfect coordination in the no noise limit of a global global game, since the potential maximizing action profile will always be played. Leister et al. (2018) report a "balancedness" condition which is sufficient for this in a class of BAS games motivated by interaction on a network.

This condition requires that payoffs are not too asymmetric across players. To see why, note that if payoffs of the game were symmetric, so $\Phi(a,\theta) = \widehat{\Phi}(n(a),\theta)$, then supermodularity implies that $\widehat{\Phi}(n+1,\theta) - \widehat{\Phi}(n,\theta)$ is increasing in n and thus (7) is satisfied. If payoffs are asymmetric, define a symmetrized potential $\widehat{\Phi}\colon \{0,\ldots,|I|\} \times \Theta \to \mathbb{R}$ by

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

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

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

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

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

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

$$\Phi(a,\theta) = \Delta(a,\theta) + \widehat{\Phi}(n,\theta) \leq \frac{n(a)}{|I|}\Phi(\mathbf{1},\theta) \tag{8}$$

and so

$$\Delta(a,\theta) \leq S(n,\theta) \tag{9}$$

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

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

$$\ell \sum_{k=1}^{|I|}(h_k - c_k) \geq |I| \sum_{k=1}^{\ell}(h_k - c_k) \tag{10}$$

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:

$$h_{k+1} - c_{k+1} \geq h_k - c_k$$

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

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

Now we have:

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

The proof is given in Appendix A.4.

4.3 Application 1: Information Design with Adversarial Equilibrium Selection

Characterizations of implementability are key ingredients in *information design* problems. Suppose that an information designer receives $V(a,\theta)$ if agents choose $a \in A$ in state $\theta \in \Theta$. We will maintain:

Assumption 1. *For each $\theta \in \Theta$, $V(a,\theta)$ is weakly increasing in a.*

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

$$\sup_{\mathcal{T}} \sum_{a,\theta} \left(\sum_t \pi(t,\theta) \prod_i \underline{\sigma}_i(t_i)(a_i) \right) V(a,\theta).$$

But under our definition of smallest equilibrium implementable outcomes, this is equivalent to

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

An optimal solution is an element of $\overline{SI}$ that maximizes $\sum_{a,\theta} \nu(a,\theta)V(a,\theta)$.

Our main result requires one additional assumption on the designer's objective:

Definition 12. 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 $\Phi(a,\theta) > \Phi(\mathbf{1},\theta)$.

Restricted convexity holds, for example, if $V(a,\theta) = \left(\frac{n(a)}{|I|}\right)^{\alpha}$ with $\alpha \geq 1$; in particular when the designer wants to maximize the expected number of players who play action 1 ($\alpha = 1$), or the probability that all players play 1 ($\alpha \to \infty$).

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

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

Theorem 2. *Suppose that Φ satisfies convexity and V satisfies restricted convexity with respect to Φ. Then there exists an optimal solution to the adversarial information design problem that satisfies the perfect coordination property.*

The proof is given in Appendix A.5.

Given Theorem 2, it is easy to characterize a solution to the information designer's problem. Since the game is uniformly convex and the solution satisfies the perfect coordination property, Proposition 4 establishes that it is enough to consider the perfect coordination outcome that maximizes the designer's objective subject to grand coalitional obedience.

To describe this outcome, we relabel the states as $\Theta = \{1, \ldots, |\Theta|\}$ in such a way that $\frac{\Phi(\mathbf{1},\theta)}{V(\mathbf{1},\theta)}$ is increasing in θ:

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

By the dominance state assumption, $\Phi(\mathbf{1},\bar{\theta}) > 0$. Under the perfect coordination property, the optimization problem boils down to

$$\max_{\nu(\mathbf{1},\cdot)} \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta) V(\mathbf{1},\theta)$$

subject to

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

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

Define

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

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

$$\begin{aligned}
\Psi(\theta^*) &= \sum_{\theta \geq \theta^*} \mu(\theta)\Phi(\mathbf{1}, \theta) \geq 0, \\
\Psi(\theta^* - 1) &= \sum_{\theta \geq \theta^* - 1} \mu(\theta)\Phi(\mathbf{1}, \theta) < 0.
\end{aligned}$$

Note that $\theta^* > 1$ and $\Phi(\mathbf{1}, \theta^* - 1) < 0$. Let

$$p^* = \frac{\Psi(\theta^*)}{-\Phi(\mathbf{1}, \theta^* - 1)}. \tag{11}$$

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

Now define the outcome ν^* by

$$\nu^*(a, \theta) = \begin{cases}
\mu(\theta) & \text{if } a = \mathbf{1} \text{ and } \theta \geq \theta^*, \\
p^* & \text{if } a = \mathbf{1} \text{ and } \theta = \theta^* - 1, \\
\mu(\theta) - p^* & \text{if } a = \mathbf{0} \text{ and } \theta = \theta^* - 1, \\
\mu(\theta) & \text{if } a = \mathbf{0} \text{ and } \theta < \theta^* - 1, \\
0 & \text{otherwise.}
\end{cases} \tag{12}$$

This satisfies the grand coalitional obedience condition with equality:

$$\sum_{(a,\theta) \in A \times \Theta} \nu^*(a, \theta)\Phi(a, \theta) = \Psi(\theta^*) + p^*\Phi(\mathbf{1}, \theta^* - 1) = 0. \tag{13}$$

Clearly, it also satisfies obedience. Its value of the objective function is

$$\begin{aligned}
\sum_{(a,\theta) \in A \times \Theta} \nu^*(a, \theta)V(a, \theta) &= \sum_{\theta \geq \theta^*} \mu(\theta)V(\mathbf{1}, \theta) + p^*V(\mathbf{1}, \theta^* - 1) \\
&= \sum_{\theta \geq \theta^*} \mu(\theta) \left(V(\mathbf{1}, \theta) + \frac{\Phi(\mathbf{1}, \theta)}{-\Phi(\mathbf{1}, \theta^* - 1)} V(\mathbf{1}, \theta^* - 1) \right).
\end{aligned}$$

Proposition 5. *Suppose that Φ satisfies convexity and V satisfies restricted convexity with respect to Φ. Then the outcome ν^* defined in (12) is an optimal solution.*

The proof is given in Appendix A.6.

The general construction in the proof of Theorem 1 provides an information structure that implements this outcome. However, because we have a potential game, a simpler argument is possible. To avoid complications, let us discuss how to smallest equilibrium implement the slightly suboptimal outcome $\hat{\nu}$ as defined by $\hat{\nu}(\mathbf{1},\theta) = \mu(\theta)$ for $\theta \geq \theta^*$ and $\hat{\nu}(\mathbf{0},\theta) = \mu(\theta)$ for $\theta < \theta^*$, while assuming that $\hat{\nu}$ satisfies coalitional obedience (hence sequential obedience):

$$\sum_{\theta \geq \theta^*} \mu(\theta)\Phi(\mathbf{1},\theta) > 0.$$

If $\theta < \theta^*$, then this fact is publicly announced. If $\theta \geq \theta^*$, then an email game/global game like signal protocol is employed: $m \in \mathbb{N}$ is drawn according to an almost uniform exponential distribution and a ranking γ is drawn according to the uniform distribution on all permutations (by symmetry), while if $\theta = |\Theta|$ (assume this is a dominance state), signals are skewed towards $1, \ldots, |I|-1$. Upon receipt of signals $1, \ldots, |I|-1$, action 1 is dominant. For signals $t_i \geq |I|$, the expected payoff to action 1 is (approximately) the expected average payoff with respect to the Laplacian belief:

$$\frac{\sum_{\ell=1}^{|I|} \sum_{\theta \geq \theta^*} \mu(\theta)\hat{d}(\ell,\theta)}{|I|} = \sum_{\theta \geq \theta^*} \mu(\theta) \frac{\hat{\Phi}(|I|,\theta) - \hat{\Phi}(0,\theta)}{|I|}.$$

This being positive is equivalent to coalitional obedience (hence sequential obedience).

4.4 Application 2: Implementing An Outcome with Information Design and Payments

In our previous application, we proved general results about the pure information design problem using our simplified characterizations of sequential obedience in Section 4.2. In this section, we demonstrate that those same results offer a simple and clean analysis to the problem of joint design of payoff and information structures in the context of contracting with externalities (Winter (2004), Moriya and Yamashita (2020)), where the designer has the ability to make payments to the players as well as designing the information.

There is a team I of agents who are engaged in a joint project. Each agent decides whether to exert effort (action 1) or not (action 0), where the effort cost is $c > 0$, common across agents. The success of the project depends on the number of agents who exert effort as well as the state of the world, unknown to the agents, which is drawn from Θ according to μ. The project's technology is given by the function $p\colon \{0,\dots,|I|\}\times\Theta \to [0,1]$, where $p(n,\theta)$ is the success probability when n agents exert effort at state θ. For $n = 1,\dots,|I|$, write

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

We assume increasing returns to scale (IRS) on p, i.e., we assume that $\mathit{\Delta} p(n,\theta)$ is strictly increasing in n for each θ.

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

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

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

The objective of the principal is to find a bonus scheme with a least total payment and an information structure that induce all types of all agents to exert effort in the smallest, hence unique, equilibrium of the induced Bayesian game. Thus, the problem becomes:

$$\inf_{(b_i)_{i\in I}:\bar\nu\in SI} \sum_{i\in I} b_i,$$

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

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

$$\mathit{\Delta} p(1,\bar\theta) \geq \sum_{\theta\in\Theta} \mu(\theta)\frac{p(|I|,\theta) - p(0,\theta)}{|I|}. \tag{14}$$

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

Proposition 6. *The unique optimal bonus scheme is given by* $(b^*, \ldots, b^*)$*, where*

$$b^* = \frac{|I|c}{\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta))}.$$

The proof, given in Appendix A.7, proceeds as follows. By Theorem 1(1), sequential obedience of $\bar{\nu}$ is a necessary condition for $\bar{\nu} \in SI$, which will give us a condition on payoffs $(d_i)_{i \in I}$, or bonuses $(b_i)_{i \in I}$. But the base game given $(b_i)_{i \in I}$ is a potential game with a potential

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

Therefore, by Proposition 3, sequential obedience reduces to the simpler condition of coalitional obedience, which is easy to inspect. It gives a lower bound of the total value of bonuses under which $\bar{\nu} \in SI$. The symmetric bonus scheme $(b^*, \ldots, b^*)$ as given in the statement attains this lower bound. Conversely, for any $\varepsilon > 0$, under $(b^* + \varepsilon, \ldots, b^* + \varepsilon)$, $\bar{\nu}$ satisfies coalitional obedience, hence sequential obedience, and also, the dominance state assumption, which is endogenous with bonus choice, is satisfied by the assumption (14): it therefore follows from Theorem 1(2) that $\bar{\nu} \in SI$. Thus, $(b^*, \ldots, b^*)$ is an optimal bonus scheme.

Let us close this section by briefly discussing related studies. The original model of Winter (2004) has no state uncertainty (i.e., $|\Theta| = 1$). Winter (2004) shows that, even with symmetric effort costs, an optimal bonus scheme must be discriminatory. Specifically, it is given by $\left(\frac{c}{\Delta p(1)}, \ldots, \frac{c}{\Delta p(|I|)}\right)$ (modulo permutation). Moriya and Yamashita (2020) introduce state uncertainty to Winter's (2004) model and study the join design of bonus payments and information allocation with two agents and two states. They derive the optimal bonus scheme restricting to symmetric schemes $b_1 = b_2$. Our analysis extends theirs to any (finite) numbers of agents and states and shows that, under the dominance state assumption (i.e., assumption (14)), a symmetric bonus scheme is indeed optimal among asymmetric schemes, and asymmetric ones are strictly suboptimal. Halac et al. (2019) also consider the join design of payments and information in Winter's (2004) model (with many agents and no state uncertainty, but with asymmetric costs), but where the payments may vary across types within an information structure (in our terminology, b_i is a function on T_i). They derive the optimal value of the total payment and the optimal "ranking scheme", which is analogous to the construction in our Theorem 1(1) (and more closely to that in Oyama and Takahashi (2020)).

5 Full Implementation

We focussed on smallest equilibrium implementation, rather than full implementation, because it is the most relevant notion for applications and is simpler to state. However, our argument easily extends to full implementation. In this case, we can show that both sequential obedience and its reverse version are necessary and jointly sufficient for full implementation.

To proceed, we add a symmetric dominant state assumption that there exists $\underline{\theta} \in \Theta$ such that $\mu(\underline{\theta}) > 0$ and $d_i(\mathbf{0}_{-i}, \underline{\theta}) > 0$ for all $i \in I$. We now interpret an ordered outcome as describing switches from action 1 to action 0. So write $\bar{a}^0(\gamma) = \mathbf{1} - \bar{a}(\gamma) \in A$ for the action profile such that player i plays action 0 if and only if i is listed in γ and $\bar{a}^0_{-i}(\gamma)$ for the action profile such that only players before i in γ play action 0. Now an ordered outcome $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$ reverse induces $\nu \in \Delta(A \times \Theta)$ if

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

Ordered outcome ν_Γ satisfies *reverse sequential obedience*

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

for all $i \in I$ such that $\nu_\Gamma(\Gamma_i \times \Theta) > 0$. An outcome ν satisfies reverse sequential obedience if there exists an ordered outcome ν_Γ that reverse induces ν and satisfies reverse sequential obedience. Finally, say that an outcome satisfies dominance if $\nu(\mathbf{0}, \underline{\theta}) > 0$ as well as $\nu(\mathbf{1}, \overline{\theta}) > 0$. Then we have

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

2. If an outcome satisfies consistency, sequential obedience, reverse sequential obedience, and dominance, then it is is fully implementable.

The proof is given in Appendix A.8.

Appendix

A.1 Proof of Theorem 1

A.1.1 Proof of Theorem 1(1)

In this proof, by abusing notation, for a pure strategy σ_i we let $\sigma_i(t_i)$ represent a pure action (an element of A_i), rather than a mixed action (an element of $\Delta(A_i)$).

Let $\nu \in \Delta(A \times \Theta)$ be smallest equilibrium implementable, and let (T, π) be a type space whose smallest Bayesian Nash equilibrium $\underline{\sigma}$ induces ν:

$$\nu(a, \theta) = \sum_{t:\underline{\sigma}(t)=a} \pi(t, \theta).$$

By Proposition 1, ν satisfies consistency and obedience.

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

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

By supermodularity, for each i and t_i, the sequence $\{\sigma_i^n(t_i)\}$ (of pure actions 0 and 1) is monotone increasing and converges to $\underline{\sigma}_i(t_i)$. Let $n_i(t_i) = n$ if $\sigma_i^{n-1}(t_i) = 0$ and $\sigma_i^n(t_i) = 1$ (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 $\gamma = (i_1, \ldots, i_k) \in \Gamma$, we say that $n(t)$ is consistent with[8] γ if $n_i(t_i) < \infty$ if and only if $i \in S(\gamma)$,[9] and for $i \in S(\gamma)$, $n_{i_\ell}(t_{i_\ell}) < n_{i_m}(t_{i_m})$ if and only if $\ell < m$. Let $T(\gamma)$ denote the set of type profiles t such that $n(t)$ is consistent with γ.

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

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

[8] Or, ordered according to.

[9] Define $S(\gamma)$ somewhere before.

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

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

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

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

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

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

Thus, ν satisfies sequential obedience.

A.1.2 Proof of Theorem 1(2)

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

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

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

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

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

$$\bar{q} d_i(\mathbf{0}_{-i}, \overline{\theta}) + (1-\bar{q}) \min_{\theta \neq \overline{\theta}} d_i(\mathbf{0}_{-i}, \theta) > 0 \tag{15}$$

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

$$\frac{\frac{\varepsilon}{|I|-1}}{\frac{\varepsilon}{|I|-1} + \eta} \geq \bar{q} \tag{16}$$

and

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I|-|a_{-i}(\gamma)|-1} \tilde{\nu}_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) > 0 \tag{17}$$

for all $i \in I$. 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

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

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 $\gamma = (i_1, \ldots, i_k) \in \Gamma$.

Claim 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 last inequality is by (16). □

Claim 2. *For any $i \in I$ and any $\tau \in \{|I|, |I|+1, \ldots\}$,*

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

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

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

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

as claimed. □

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

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

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

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

We are in a position to conclude the proof of Theorem 1. We first show that action 1 is uniquely rationalizable for all players of types $t_i < \infty$. For types $t_i \leq |I|-1$, action 1 is a strictly dominant action by Claim 1 and condition (15). For $\tau \geq |I|$, suppose that action 1 is uniquely rationalizable for all players of types $t_i \leq \tau - 1$. Then the expected payoff for a player i of type $t_i = \tau$ from playing action 1 is no smaller than

$$\sum_{S \subset I \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \tau\} = S, \theta | t_i = \tau) d_i(\mathbf{1}_S, \theta)$$

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

where the equality is by Claim 2 and the inequality by the "perturbed" sequential obedience condition (17). 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 $\underline{\sigma}_i(t_i)(1) = 1$ 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 given by

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

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

A.1.3 Proof of Corollary 1

The "only if" part is easy to prove. To prove the "if" part, let $\nu \in \Delta(A \times \Theta)$ satisfy consistency, obedience, and weak sequential obedience with $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$. Let $\bar{\gamma} \in \Gamma$ be a permutation of all players, say, $\bar{\gamma} = 1 \cdots |I|$. For $\varepsilon > 0$, define $\nu_\Gamma^\varepsilon \in \Delta(\Gamma \times \Theta)$ by

$$\nu_\Gamma^\varepsilon(\gamma, \theta) = \begin{cases} (1-\varepsilon)\nu_\Gamma(\gamma, \theta) + \varepsilon & \text{if } [\gamma = \bar{\gamma} \text{ and } \theta = \bar{\theta}] \text{ or } [\gamma = \emptyset \text{ and } \theta \neq \bar{\theta}], \\ (1-\varepsilon)\nu_\Gamma(\gamma, \theta) & \text{otherwise,} \end{cases}$$

and $\nu^\varepsilon \in \Delta(A \times \Theta)$ by $\nu^\varepsilon(a, \theta) = \sum_{\gamma : \bar{a}(\gamma) = a} \nu_\Gamma^\varepsilon(\gamma, \theta)$, which satisfies consistency and dominance. The outcome ν^ε satisfies 0-obedience since for all $i \in I$, we have

$$\begin{aligned} &\sum_{a_{-i}, \theta} \nu^\varepsilon((0, a_{-i}), \theta) d_i(a_{-i}, \theta) \\ &= \sum_{a_{-i}, \theta} \nu((0, a_{-i}), \theta) d_i(a_{-i}, \theta) - \varepsilon \sum_{a_{-i}} \nu((0, a_{-i}), \bar{\theta}) d_i(a_{-i}, \bar{\theta}) \\ &\quad - \varepsilon \sum_{a_{-i}, \theta \neq \bar{\theta}} \nu((0, a_{-i}), \theta)(d_i(a_{-i}, \theta) - d_i(\mathbf{0}_{-i}, \theta)) \\ &\leq \sum_{a_{-i}, \theta} \nu((0, a_{-i}), \theta) d_i(a_{-i}, \theta) \leq 0, \end{aligned}$$

where the first inequality follows from dominance state and monotonicity. It also satisfies sequential obedience since for all $i \in I$ such that $\nu_\Gamma^\varepsilon(\Gamma_i \times \Theta) > 0$, for which $\nu_\Gamma(\Gamma_i \times \Theta) > 0$, we have

$$
\begin{aligned}
&\sum_{\gamma \in \Gamma_i, \theta} \nu_\Gamma^\varepsilon(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) \\
&= (1-\varepsilon) \sum_{\gamma \in \Gamma_i, \theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) + \varepsilon d_i(a_{-i}(\bar{\gamma}), \bar{\theta}) \\
&> (1-\varepsilon) \sum_{\gamma \in \Gamma_i, \theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta) \geq 0,
\end{aligned}
$$

where the strict inequality follows from dominance state. Hence, we have $\nu^\varepsilon \in SI$ by Theorem 1. Since $\nu^\varepsilon \to \nu$ as $\varepsilon \to 0$, we therefore have $\nu \in \overline{SI}$.

A.1.4 Proof of Corollary 2

Let ν satisfy consistency and weak sequential obedience. By (the proof of) Corollary 1, there exists a sequence (ν^n) of outcomes converging to ν that satisfy consistency, sequential obedience, and dominance. For each n, the smallest equilibrium of the type space as constructed in the proof of Theorem 1(2) induces an outcome $\tilde{\nu}^n$ that first-order stochastically dominates ν while satisfying consistency, obedience, and sequential obedience. Then a limit point of $(\tilde{\nu}^n)$ first-order stochastically dominates ν and satisfies consistency, obedience, and weak sequential obedience.

A.2 Proof of Proposition 2

The "only if" part: Suppose that $\nu \in \Delta(A \times \Theta)$ satisfies sequential obedience (resp. weak sequential obedience) with $\nu_\Gamma \in \Delta(\Gamma \times \Theta)$. Then for any $(\lambda_i)_{i \in S(\nu)} \geq 0$, $(\lambda_i)_{i \in S(\nu)} \neq 0$, we have

$$
\begin{aligned}
&\sum_{a \in A(\nu), \theta \in \Theta} \nu(a, \theta) \max_{\gamma \in \Gamma(a)} \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) \\
&= \sum_{a \in A(\nu), \theta \in \Theta} \sum_{\gamma' \in \Gamma(a)} \nu_\Gamma(\gamma', \theta) \max_{\gamma \in \Gamma(a)} \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) \\
&\geq \sum_{a \in A(\nu), \theta \in \Theta} \sum_{\gamma \in \Gamma(a)} \nu_\Gamma(\gamma, \theta) \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) \\
&= \sum_{i \in S(\nu)} \lambda_i \sum_{\gamma \in \Gamma_i, \theta \in \Theta} \nu_\Gamma(\gamma, \theta) d_i(a_{-i}(\gamma), \theta),
\end{aligned}
$$

which is positive (resp. nonnegative) by sequential obedience (resp. weak sequential obedience).

The "if" part: Suppose that (4) holds. Let $\Gamma(\nu) = \{\gamma \in \Gamma \mid S(\gamma) \subset S(\nu)\}$, and $N(\nu) = \{\nu_\Gamma \in \Delta(\Gamma(\nu) \times \Theta) \mid \sum_{\gamma:\bar{a}(\gamma)=a} \nu_\Gamma(\gamma, \theta) = \nu(a, \theta)\}$, which is convex and compact. For $\nu_\Gamma \in N(\nu)$ and $\lambda \in \Delta(S(\nu))$, let

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

which is linear in each of ν_Γ and λ. Note that ν satisfies sequential obedience (resp. weak sequential obedience) if and only if there exists $\bar{\nu}_\Gamma \in N(\nu)$ ($\subset \Delta(\Gamma(\nu) \times \Theta)$ such that $D(\bar{\nu}_\Gamma, \lambda) >$ (resp. $\geq$) 0 for all $\lambda \in \Delta(S(\nu))$, where such a $\bar{\nu}_\Gamma$ is naturally extended to $\Delta(\Gamma \times \Theta)$ by $\bar{\nu}_\Gamma(\gamma, \theta) = 0$ for $\gamma \notin \Gamma(\nu)$.

Now, by the minimax theorem, there exist $\bar{\nu}_\Gamma \in N(\nu)$ and $\bar{\lambda} \in \Delta(S(\nu))$ such that

$$D(\nu_\Gamma, \bar{\lambda}) \leq D(\bar{\nu}_\Gamma, \bar{\lambda}) \leq D(\bar{\nu}_\Gamma, \lambda)$$

for all $\nu_\Gamma \in N(\nu)$ and $\lambda \in \Delta(S(\nu))$. It therefore suffices to show that $D(\bar{\nu}_\Gamma, \bar{\lambda}) >$ (resp. $\geq$) 0.

For each $(a, \theta) \in A(\nu) \times \Theta$, let $\gamma_{a,\theta} \in \Gamma(a)$ be a sequence that maximizes $\sum_{i \in S(\gamma)} \bar{\lambda}_i d_i(a_{-i}(\gamma), \theta)$. Define $\nu'_\Gamma \in N(\nu)$ by $\nu'_\Gamma(\gamma, \theta) = \nu(a, \theta)$ if $\gamma = \gamma_{a,\theta}$ and $\nu'_\Gamma(\gamma, \theta) = 0$ otherwise. Then we have

$$\begin{aligned} D(\bar{\nu}_\Gamma, \bar{\lambda}) &\geq D(\nu'_\Gamma, \bar{\lambda}) \\ &= \sum_{a \in A(\nu), \theta \in \Theta} \nu(a, \theta) \sum_{i \in S(\gamma_{a,\theta})} \bar{\lambda}_i d_i(a_{-i}(\gamma_{a,\theta}), \theta), \end{aligned}$$

which is positive (resp. nonnegative) by (4), as desired.

A.3 Proof of Proposition 3

Suppose that the game admits a potential Φ. By Proposition 2, it suffices to show that $\nu \in \Delta(A \times \Theta)$ satisfies condition (4) in Proposition 2 if and only if it satisfies coalitional obedience (resp. weak coalitional obedience). Recall $S(\nu) = \{i \in I \mid \nu((1, a_{-i}), \theta) > 0$ for some $a_{-i} \in A_{-i}$ and $\theta \in \Theta\}$ and $S(a) = \{i \in I \mid a_i = 1\}$ for $a \in A$. Note that a dominates ν if and only if $S(\nu) \subset S(a)$, and in particular, $\mathbf{1}_{S(\nu)}$ dominates ν, so that $\Phi_\nu(\mathbf{1}) = \Phi_\nu(\mathbf{1}_{S(\nu)})$.

The "only if" part: Suppose that ν satisfies sequential obedience (resp. weak sequential obedience) and hence condition (4). Fix any $a \in A$ that does not dominate ν. Define $(\lambda_i)_{i \in S(\nu)} \geq 0$ by $\lambda_i = 1$ if $i \in S(\nu) \setminus S(a)$ $(\neq \emptyset)$ and $\lambda_i = 0$ if $i \in S(\nu) \cap S(a)$.

Consider any $(a', \theta) \in A(\nu) \times \Theta$, where $A(\nu) = \{a \in A \mid S(a) \subset S(\nu)\}$. By supermodularity, any sequence $\gamma \in \Gamma(a')$ that maximizes $\sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) = \sum_{i \in S(a') \setminus S(a)} d_i(a_{-i}(\gamma), \theta)$ ranks all players in $S(a') \cap S(a)$ earlier than those in $S(a') \setminus S(a)$. Thus,

$$\max_{\gamma \in \Gamma(a')} \sum_{i \in S(\gamma)} \lambda_i d_i(a_{-i}(\gamma), \theta) = \Phi(a', \theta) - \Phi(a \wedge a', \theta).$$

Therefore, we have

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

which is positive (resp. nonnegative) by condition (4).

The "if" part: Suppose that ν satisfies coalitional obedience (resp. weak coalitional obedience), so that $\Phi_\nu(\mathbf{1}) = \Phi_\nu(\mathbf{1}_{S(\nu)}) >$ (resp. $\geq$) $\Phi_\nu(a)$ for all a that does not dominate ν. We want to show that ν satisfies condition (4). Fix any $(\lambda_i)_{i \in S(\nu)} \geq 0$, $(\lambda_i)_{i \in S(\nu)} \neq 0$. Let $\gamma = (i_1, \ldots, i_{|S(\nu)|}) \in \Gamma(S(\nu))$ be a permutation of players in $S(\nu)$ such that $\lambda_{i_1} \leq \cdots \leq \lambda_{i_{|S(\nu)|}}$. Then we have

$$\begin{aligned}
&\text{LHS of (4)} \\
&\geq \sum_{a' \in A(\nu), \theta \in \Theta} \nu(a', \theta) \sum_{i \in S(a') \cap S(\nu)} \lambda_i(\Phi((1, a_{-i}(\gamma_\lambda)), \theta) - \Phi((0, a_{-i}(\gamma_\lambda)), \theta)) \\
&= \sum_{a' \in A(\nu), \theta \in \Theta} \nu(a', \theta) \sum_{i \in S(\nu)} \lambda_i(\Phi((1, a_{-i}(\gamma_\lambda)) \wedge a', \theta) - \Phi((0, a_{-i}(\gamma_\lambda)) \wedge a'), \theta)) \\
&= \sum_{i \in S(\nu)} \lambda_i \sum_{a' \in A(\nu), \theta \in \Theta} \nu(a', \theta)(\Phi((1, a_{-i}(\gamma_\lambda)) \wedge a', \theta) - \Phi((0, a_{-i}(\gamma_\lambda)) \wedge a'), \theta)) \\
&= \sum_{i \in S(\nu)} \lambda_i(\Phi_\nu(1, a_{-i}(\gamma_\lambda)) - \Phi_\nu(0, a_{-i}(\gamma_\lambda))) \\
&= \sum_{k=1}^{|S(\nu)|} (\lambda_{i_k} - \lambda_{i_{k-1}}) \sum_{\ell=k}^{|S(\nu)|} (\Phi_\nu(1, a_{-i_\ell}(\gamma_\lambda)) - \Phi_\nu(0, a_{-i_\ell}(\gamma_\lambda))) \\
&= \sum_{k=1}^{|S(\nu)|} (\lambda_{i_k} - \lambda_{i_{k-1}})(\Phi_\nu(\mathbf{1}_{S(\nu)}) - \Phi_\nu(\mathbf{1}_{\{i_1, \ldots, i_{k-1}\}})),
\end{aligned}$$

which is positive (resp. nonnegative) by coalitional obedience (resp. weak coalitional obedience), as desired, where $\mathbf{1}_{\{i_1,\ldots,i_{k-1}\}}$ does not dominate ν.

A.4 Proof of Proposition 4

By Proposition 3, sequential obedience (resp. weak sequential obedience) is equivalent to weak coalitional obedience (resp. weak coalitional obedience) in a potential game. The "only if" part is obvious. The "if" direction follows from convexity of Φ since for a perfect coordination outcome ν, we have

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

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

A.5 Proof of Theorem 2

Suppose that Φ satisfies convexity and V satisfies restricted convexity with respect to Φ. For each (a,θ), let $\alpha(a,\theta) \in [0,1]$ by

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

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

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

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

which satisfies the perfect coordination property. Since ν is consistent with μ, so is ν'. Since ν satisfies weak coalitional obedience, we also have

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

Therefore, ν' satisfies weak grand coalitional obedience. Hence, by the convexity of Φ, we have $\nu' \in \overline{SI}$ by Proposition 4 and Corollary 1.

For the value of the objective function, we have

$$\begin{aligned}\sum_{a\in A,\theta\in\Theta} \nu'(a,\theta)V(a,\theta) &= \sum_{\theta\in\Theta} [\nu'(\mathbf{0},\theta)V(\mathbf{0},\theta) + \nu'(\mathbf{1},\theta)V(\mathbf{1},\theta)] \\ &= \sum_{a\in A,\theta\in\Theta} \nu(a,\theta)\left[(1-\alpha(a,\theta))V(\mathbf{0},\theta) + \alpha(a,\theta)V(\mathbf{1},\theta)\right] \\ &\geq \sum_{a\in A,\theta\in\Theta} \nu(a,\theta)V(a,\theta).\end{aligned}$$

Thus, an optimal outcome exists among those that satisfy the perfect coordination property.

A.6 Proof of Proposition 5

Let $\nu(\mathbf{1},\cdot)$ be such that $0 \leq \nu(\mathbf{1},\theta) \leq \mu(\theta)$ and $\sum_{\theta\in\Theta} \nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta) < \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)V(\mathbf{1},\theta)$. Define $\xi(\cdot)$, $\xi^*(\cdot)$, and $\xi^{**}(\cdot)$ 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 $\xi^{**}(\theta^*-1) = \xi^*(\theta^*-1) + \sum_{\theta\in\Theta} \nu(\mathbf{1},\theta)V(\mathbf{1},\theta) - \sum_{\theta\in\Theta} \nu^*(\mathbf{1},\theta)V(\mathbf{1},\theta) > \xi^*(\theta^*-1)$ and $\xi^{**}(\theta) = \xi^*(\theta)$ for all $\theta \neq \theta^*-1$.

Since $\xi^{**}(\cdot)$ first-order stochastically dominates $\xi(\cdot)$ and $\frac{\Phi(\mathbf{1},\theta)}{V(\mathbf{1},\theta)}$ is increasing in θ, we have

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

But we have

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

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

since the first term equals 0 by (13), and $\Phi(\mathbf{1}, \theta^* - 1) < 0$. This means that $\nu(\mathbf{1}, \cdot)$ is not feasible.

A.7 Proof of Proposition 6

Since the base game given $(b_i)_{i \in I}$ is a potential game with a potential

$$\Phi(a, \theta) = p(n(a), \theta) - p(0, \theta) - \sum_{i \in I} a_i \frac{c}{b_i},$$

sequential obedience is equivalent to coalitional obedience: $\Phi_{\bar{\nu}}(\mathbf{1}) > \Phi_{\bar{\nu}}(a)$ for all $a \neq \mathbf{1}$ (Proposition 3). A necessary condition is thus $\Phi_{\bar{\nu}}(\mathbf{1}) > \Phi_{\bar{\nu}}(\mathbf{0})$, which is written as

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

or

$$\sum_{i \in I} \frac{1}{b_i} < \frac{\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta))}{c}.$$

By the strict convexity of the function $x \mapsto 1/x$, we have $1/\sum_i b_i/|I| \leq (\sum_i (1/b_i))/|I|$, or $|I|^2/\sum_i b_i \leq \sum_i (1/b_i)$, where a strict inequality holds unless b_i's are identical. Hence, by Theorem 1(1), if a scheme $(b_i)_{i \in I}$ smallest equilibrium implements $\bar{\nu}$, it is necessary that

$$\frac{|I|^2}{\sum_{i \in I} b_i} < \frac{\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta))}{c},$$

or

$$\sum_{i \in I} b_i > \frac{|I|^2 c}{\sum_{\theta \in \Theta} \mu(\theta)(p(|I|, \theta) - p(0, \theta))}.$$

Thus, the right hand side is a lower bound of the total bonus payment for bonus schemes under which $\bar{\nu} \in SI$. As discussed above, by the strict convexity of $x \mapsto 1/x$, an optimal scheme must entail symmetric payments.

Now let b^* be as given in the statement. The total bonus $|I|b^*$ is in fact equal to the lower bound obtained above. We want to show that $\bar{\nu} \in SI$ under $(b^* + \varepsilon, \ldots, b^* + \varepsilon)$ for every $\varepsilon > 0$. The potential function is now

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

which is convex. Therefore, sequential obedience is equivalent to grand coalitional obedience: $\Phi_{\bar{\nu}}(\mathbf{1}) > \Phi_{\bar{\nu}}(\mathbf{0})$ (Proposition 4), which is written as

$$\sum_{\theta\in\Theta} \mu(\theta)(p(|I|,\theta) - p(0,\theta)) - |I|\frac{c}{b^* + \varepsilon} > 0.$$

This is satisfied for every $\varepsilon > 0$ by the definition of b^*. Finally, by the assumption (14), we have $b^* \geq c/\Delta p(1,\bar{\theta})$, and therefore,

$$d_i(\mathbf{0}_{-i},\bar{\theta}) = \Delta p(1,\bar{\theta}) - \frac{c}{b^* + \varepsilon} > 0,$$

so that the dominance state assumption is satisfied for any $\varepsilon > 0$. Hence, by Theorem 1(2), we have $\bar{\nu} \in SI$ for any $\varepsilon > 0$.

A.8 Proof of Theorem 3

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

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

and

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

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

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

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

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

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

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

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

By richness, we can take a $\bar{q} < 1$ such that

$$\bar{q} d_i(\mathbf{0}_{-i}, \overline{\theta}) + (1-\bar{q}) \min_{\theta \neq \overline{\theta}} d_i(\mathbf{0}_{-i},\theta) > 0, \tag{18}$$

$$\bar{q} d_i(\mathbf{1}_{-i}, \underline{\theta}) + (1-\bar{q}) \max_{\theta \neq \underline{\theta}} d_i(\mathbf{1}_{-i},\theta) < 0 \tag{19}$$

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

$$\frac{\frac{\varepsilon}{|I|-1}}{\frac{\varepsilon}{|I|-1} + \eta} \geq \bar{q}, \tag{20}$$

and

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I| - |a_{-i}(\gamma)| - 1} \tilde{\nu}_{\Gamma}^{+}(\gamma,\theta) d_i(a_{-i}(\gamma),\theta) > 0 \tag{21}$$

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

$$\sum_{\gamma \in \Gamma_i, \theta \in \Theta} (1-\eta)^{|I| - |a^0_{-i}(\gamma)| - 1} \tilde{\nu}_{\Gamma}^{-}(\gamma,\theta) d_i(a^0_{-i}(\gamma),\theta) < 0 \tag{22}$$

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

Now construct the type space (T,π) as follows. For each $i \in I$, let $T_i = \{1,2,\ldots\} \times \{0,1\}$. Let $\pi \in \Delta(T \times \Theta)$ be given by

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

for each $t = (s_i,a_i)_{i \in I} \in T$ and $\theta \in \Theta$, where for $\gamma = (i_1,\ldots,i_k) \in \Gamma$ and $i \in S(\gamma)$, $\ell(i,\gamma) = \ell$ if $i = i_\ell$. Observe that π is consistent with μ: $\sum_t \pi(t,\theta) = \mu(\theta)$ for all $\theta \in \Theta$.

Claim 4. *For any $i \in I$ and any $\tau \in \{1, \ldots, |I|-1\}$, $\pi(\overline{\theta}|t_i = (\tau, 1)) \geq \bar{q}$ and $\pi(\underline{\theta}|t_i = (\tau, 0)) \geq \bar{q}$.*

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

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

where the last inequality is by (20). A symmetric argument verifies the other claim. □

Claim 5. *For any $i \in I$ and any $\tau \in \{|I|, |I| + 1, \ldots\}$,*

$$\begin{aligned}
&\pi(\{j \neq i \mid s_j < \tau, a_j = 1\} = S, \theta|t_i = (\tau, 1)) \\
&\quad = (1 - \eta)^{|I|-|S|-1} \tilde{\nu}_\Gamma^+(\{\gamma^+ \in \Gamma_i \mid a_{-i}(\gamma^+) = \mathbf{1}_S\} \times \{\theta\})/C_i^+, \\
&\pi(\{j \neq i \mid s_j < \tau, a_j = 0\} = S, \theta|t_i = (\tau, 0)) \\
&\quad = (1 - \eta)^{|I|-|S|-1} \tilde{\nu}_\Gamma^-(\{\gamma^- \in \Gamma_i \mid a_{-i}^0(\gamma^-) = \mathbf{0}_S\} \times \{\theta\})/C_i^-
\end{aligned}$$

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

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

$$\begin{aligned}
&\pi(\{j \neq i \mid s_j < \tau, a_j = 1\} = S, \theta|t_i = (\tau, 1)) \\
&\quad = \pi(t_i = (\tau, 1), \{j \neq i \mid s_j < \tau, a_j = 1\} = S, \theta)/\pi(t_i = (\tau, 1)) \\
&\quad = (1 - 2\varepsilon)\eta(1 - \eta)^{\tau-|S|-1} \tilde{\nu}_\Gamma^+(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = \mathbf{1}_S\} \times \{\theta\})/\pi(t_i = (\tau, 1)) \\
&\quad = (1 - \eta)^{|I|-|S|-1} \tilde{\nu}_\Gamma^+(\{\gamma \in \Gamma_i \mid a_{-i}(\gamma) = \mathbf{1}_S\} \times \{\theta\})/C_i^+,
\end{aligned}$$

as claimed. A symmetric argument verifies the other claim. □

We are in a position to conclude the proof of Theorem 3(2). We show that action 1 is uniquely rationalizable for all players of types $t_i = (s_i, a_i)$ with $a_i = 1$. First, for types $t_i = (s_i, 1)$ with $s_i \leq |I| - 1$, action 1 is a strictly dominant action by Claim 4 and condition (18). Then, for $\tau \geq |I|$, suppose that action 1 is uniquely rationalizable for all players of types $t_i = (s_i, 1)$ with $s_i \leq \tau - 1$. Then the expected payoff for a player i of type $t_i = (\tau, 1)$ from playing action 1 is no smaller than

$$\sum_{S \subset I \setminus \{i\}, \theta \in \Theta} \pi(\{j \neq i \mid t_j < \tau\} = S, \theta|t_i = (\tau, 1)) d_i(\mathbf{1}_S, \theta)$$

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

where the equality is by Claim 5 and the inequality by the "perturbed" sequential obedience condition (21). Therefore, action 1 is uniquely rationalizable for $t_i = \tau$. Hence, by induction, action 1 is uniquely rationalizable for all types $t_i = (s_i, a_i)$ with $a_i = 1$. A symmetric argument shows that action 0 is uniquely rationalizable for all players of types $t_i = (s_i, a_i)$ with $a_i = 0$.

Finally, by construction, the unique rationalizable strategy profile, hence the unique equilibrium, induces ν, as desired.

References

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

Arieli, I., Y. Babichenko, F. Sandomirskiy, and O. Tamuz (2020): "Feasible Joint Posterior Beliefs," Tech. rep., Technion, Haifa and CalTech.

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

——— (1987): "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55, 1–18.

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

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

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

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

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

Doval, L. and J. Ely (2019): "Sequential Information Design," Tech. rep.

Feinberg, Y. (2000): "Characterizing Common Priors in the Form of Posteriors," *Journal of Economic Theory*, 91, 127–179.

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

Goldstein, I. and C. Huang (2016): "Bayesian Persuasion in Coordination Games," *American Economic Review Papers and Proceedings*, 106, 592–596.

Halac, M., E. Lipnowski, and D. Rappoport (2019): "Rank Uncertainty in Organizations," Tech. rep., Yale University, Columbia University and University of Chicago.

Hoshino, T. (2018): "Multi-Agent Persuasion: Leveraging Strategic Uncertainty," Tech. rep.

Inostroza, N. and A. Pavan (2019): "Persuasion in Global Games," Tech. rep.

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

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

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2.

Leister, C. M., Y. Zenou, and J. Zhou (2018): "Coordination on Networks," Tech. rep.

Li, F., Y. Song, and M. Zhao (2019): "Global Manipulation by Local Obfuscation: Information Design in Coordination Games," Tech. rep.

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

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

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

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

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

Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62, 1327–1347.

Morris, S. and H. S. Shin (2007): "Common Belief Foundations of Global Games," Mimeo.

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

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

——— (2019): "Incomplete Information Robustness," Tech. rep., Massachusetts Institute of Technology and Hitotsubashi University.

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

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

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

Samet, D. (1998): "Common Priors and the Separation of Convex Sets," *Games and Economic Behavior*, 24, 131–141.

Sandmann, C. (2020): "Recursive Information Design," Tech. rep., Toulouse School of Economics.

Segal, I. (1999): "Contracting with Externalities," *Quarterly Journal of Economics*, 144, 337–388.

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

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

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

Ziegler, G. (2020): "Adversarial Bilateral Information Design," Tech. rep., Northwestern University.