ROBUSTNESS IN BINARY-ACTION SUPERMODULAR GAMES REVISITED

DAISUKE OYAMA AND SATORU TAKAHASHI

ABSTRACT. We show that in all (whether generic or nongeneric) binary-action supermodular games, an extreme action profile is robust to incomplete information if and only if it is a monotone potential maximizer. The equivalence does not hold for nonextreme action profiles.

1. INTRODUCTION

In Kajii and Morris (1997), a Nash equilibrium a^* of a complete information game $\mathbf{g}$ is said to be *robust to incomplete information* if in any incomplete information game where with high probability, all players know that their payoff functions are given by those in $\mathbf{g}$, there exists a Bayesian Nash equilibrium that plays a^* with high probability. The literature has provided several sufficient conditions for robustness. In particular, Morris and Ui (2005) showed that a *monotone potential maximizer* is robust in many-action supermodular games (and in games with a supermodular monotone potential).[1]

In Oyama and Takahashi (2020), we established the necessity of monotone potential maximization for robustness in *generic* binary-action supermodular games. More precisely, we employed a strict version of monotone potential maximizer (*strict monotone potential maximizer*) and showed by contrapositive that in a binary-action supermodular game, if an action profile a^* is not a strict monotone potential maximizer, then under a genericity assumption, it is not robust, i.e., there are incomplete information perturbations where the behavior of any equilibrium is bounded away from a^*. The proof was

(D. Oyama) FACULTY OF ECONOMICS, UNIVERSITY OF TOKYO
(S. Takahashi) FACULTY OF ECONOMICS, UNIVERSITY OF TOKYO; DEPARTMENT OF ECONOMICS, NATIONAL UNIVERSITY OF SINGAPORE
E-mail addresses: oyama@e.u-tokyo.ac.jp, satorut@e.u-tokyo.ac.jp.
Date: February 22, 2024; this version: June 26, 2024.
Keywords. Incomplete information, supermodular game, robustness, monotone potential, sequential obedience.
We thank the anonymous referees for the comments. Daisuke Oyama gratefully acknowledges financial support from JSPS KAKENHI Grants 18KK0359, 19K01556, and 23K01303. Satoru Takahashi gratefully acknowledges financial support from JSPS KAKENHI Grant 23K01303.

[1]See Kajii and Morris (2020) for a brief survey of the literature.

based on a duality argument, which proceeded as follows. First, we gave a dual characterization of the nonexistence of a strict monotone potential. It is a system of certain *weak* inequalities with respect to a probability distribution over sequences of players, which can be interpreted as an obedience condition—later termed "sequential obedience" in Morris et al. (2024). Then, assuming, by genericity, *strict* inequalities in that system, we constructed the desired incomplete information perturbations such that in any equilibrium, players have strict incentives to play actions different from those in the action profile that we wanted to prove not to be robust. The result, however, does not hold in some nongeneric games: for example, in games with constant payoff functions, no action profile is a strict monotone potential maximizer, while all action profiles are robust (and monotone potential maximizers).

The present paper is to pursue the logical relationship between robustness and monotone potential maximization without relying on any extra genericity assumption. We show that in *all* binary-action supermodular games, if an extreme action profile, one where all players play action 0 or the other where all play action 1, is robust, then it is a monotone potential maximizer. Thus, combined with the result by Morris and Ui (2005), the two conditions are equivalent for extreme action profiles in these games. The proof strategy is the same as that of Oyama and Takahashi (2020), but here we take an extra care of complications arising from possible payoff ties. For nonextreme action profiles, on the other hand, the equivalence does not hold: we report an example (from Oyama and Takahashi (2019)) where a nonextreme action profile is robust but not a monotone potential maximizer.

In the final section, we also discuss two alternative versions of robustness, *approximate robustness* (Haimanko and Kajii (2016)) and *strict robustness* (Morris et al. (2023)). In contrast to the original version of Kajii and Morris (1997), these admit characterizations, in terms of strict monotone potential maximizer, that apply to all (whether extreme or nonextreme) action profiles in all (whether generic or nongeneric) binary-action supermodular games.

2. Framework

2.1. **Robustness.** A complete information game consists of a finite set I of players ($|I| \geq 2$), a finite set A_i of actions for each $i \in I$, and a payoff function $g_i \colon A \to \mathbb{R}$ for each

$i \in I$, where we denote $A = \prod_{i \in I} A_i$ and $A_{-i} = \prod_{j \neq i} A_j$ as usual. We refer to a complete information game by the profile $\mathbf{g} = (g_i)_{i \in I}$ of its payoff functions.

An elaboration of a complete information game $\mathbf{g}$ is an incomplete information game consisting of the same sets of players and actions, a countable set T_i of each player i's types, a common prior $P \in \Delta(T)$, where $T = \prod_{i \in I} T_i$, and a profile $\mathbf{u} = (u_i)_{i \in I}$ of bounded payoff functions $u_i \colon A \times T \to \mathbb{R}$.[2] We assume that $P(\{t_i\} \times T_{-i}) > 0$ for each $i \in I$ and $t_i \in T_i$, where $T_{-i} = \prod_{j \neq i} T_j$. We refer to an elaboration as $(T, P, \mathbf{u})$. For $\eta \geq 0$, a profile $\sigma = (\sigma_i)_{i \in I}$ of behavioral strategies $\sigma_i \colon T_i \to \Delta(A_i)$ is an (interim) η-Bayesian Nash equilibrium of $(T, P, \mathbf{u})$ if for all $i \in I$, $t_i \in T_i$, and $a_i, a_i' \in A_i$,

$$\begin{aligned} &\sigma_i(a_i|t_i) > 0 \\ &\quad \Longrightarrow \sum_{t_{-i} \in T_{-i}} P(t_{-i}|t_i)(u_i((a_i, \sigma_{-i}(t_{-i})), (t_i, t_{-i})) - u_i((a_i', \sigma_{-i}(t_{-i})), (t_i, t_{-i}))) \geq -\eta, \end{aligned}$$

where $P(t_{-i}|t_i) = P(t_i, t_{-i})/P(\{t_i\} \times T_{-i})$, $\sigma_{-i}(t_{-i}) = (\sigma_j(t_j))_{j \neq i}$, and the domain of u_i is extended to mixed action profiles in the standard way. A Bayesian Nash equilibrium of $(T, P, \mathbf{u})$ is a 0-Bayesian Nash equilibrium of $(T, P, \mathbf{u})$.

Given an elaboration $(T, P, \mathbf{u})$ of $\mathbf{g}$, let $T_i^{g_i}$ denote the set of player i's types that know that the payoffs are given by g_i:

$$T_i^{g_i} = \{t_i \in T_i \mid u_i(a, (t_i, t_{-i})) = g_i(a) \text{ for all } a \in A \text{ and } t_{-i} \in T_{-i} \text{ with } P(t_i, t_{-i}) > 0\},$$

and write $T^{\mathbf{g}} = \prod_{i \in I} T_i^{g_i}$. We say that $(T, P, \mathbf{u})$ is an *ε-elaboration* of $\mathbf{g}$ if $P(T^{\mathbf{g}}) \geq 1 - \varepsilon$.

Definition 1. An action profile $a^* \in A$ is *robust* in $\mathbf{g}$ if for every $\delta > 0$, there exists $\varepsilon > 0$ such that every ε-elaboration of $\mathbf{g}$ has a Bayesian Nash equilibrium σ such that $\sum_{t \in T} P(t) \prod_{i \in I} \sigma_i(a_i^*|t_i) \geq 1 - \delta$.

In what follows, we assume that the complete information game $\mathbf{g}$ has binary actions and supermodular payoff functions, i.e., for each $i \in I$, $A_i = \{0, 1\}$ and the payoff increment

$$f_i(a_{-i}) := g_i(1, a_{-i}) - g_i(0, a_{-i})$$

is weakly increasing in $a_{-i} \in A_{-i}$. For $S \subset I$, we write $A_S = \prod_{i \in S} A_i$ and let $\mathbf{0}_S \in A_S$ (resp. $\mathbf{1}_S \in A_S$) represent the action profile of players in S where all these players play action 0 (resp. 1). By convention, we write $\mathbf{0} = \mathbf{0}_I$, $\mathbf{1} = \mathbf{1}_I$, and $\mathbf{1}_{-i} = \mathbf{1}_{I \setminus \{i\}}$.

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

2.2. **(Strict) Monotone Potential Maximizers.** Oyama and Takahashi (2020) employed the strict version of monotone potential maximizer due to Oyama et al. (2008), which in binary-action games is equivalently defined as follows:[3]

Definition 2. For a binary-action game $\mathbf{g}$, an action profile $a^* \in A$ is a *strict monotone potential maximizer* in $\mathbf{g}$ if there exist a function $v\colon A \to \mathbb{R}$ and $(\lambda_i)_{i\in I}$ with $\lambda_i > 0$ for all $i \in I$ such that

$$\lambda_i f_i(a_{-i}) \geq v(1, a_{-i}) - v(0, a_{-i})$$

for all $i \in I$ such that $a_i^* = 1$ and all $a_{-i} \in A_{-i}$,

$$\lambda_i f_i(a_{-i}) \leq v(1, a_{-i}) - v(0, a_{-i})$$

for all $i \in I$ such that $a_i^* = 0$ and all $a_{-i} \in A_{-i}$, and $v(a^*) > v(a)$ for all $a \in A \setminus \{a^*\}$.

Such a function v is called a *strict monotone potential* for a^* in $\mathbf{g}$.

By definition, a strict monotone potential maximizer is a strict Nash equilibrium.

Define

$$I^1 = \{i \in I \mid f_i(a_{-i}) > 0 \text{ for some } a_{-i} \in A_{-i}\},$$
$$I^0 = \{i \in I \mid f_i(a_{-i}) < 0 \text{ for some } a_{-i} \in A_{-i}\}.$$

By definition, action 0 (resp. 1) is a weakly dominant action for players in $I^0 \setminus I^1$ (resp. $I^1 \setminus I^0$), and $f_i \equiv 0$ for players $i \in I \setminus (I^1 \cup I^0)$. Morris and Ui's (2005) (nonstrict) version of monotone potential maximizer can be equivalently defined as follows:[4]

Definition 3. For a binary-action game $\mathbf{g}$, an action profile $a^* \in A$ is a *monotone potential maximizer* in $\mathbf{g}$ if there exist a function $v\colon A \to \mathbb{R}$ and $(\lambda_i)_{i\in I}$ with $\lambda_i > 0$ for all $i \in I$ such that

$$\lambda_i f_i(a_{-i}) \geq v(1, a_{-i}) - v(0, a_{-i})$$

for all $i \in I^0$ such that $a_i^* = 1$ and all $a_{-i} \in A_{-i}$,

$$\lambda_i f_i(a_{-i}) \leq v(1, a_{-i}) - v(0, a_{-i})$$

for all $i \in I^1$ such that $a_i^* = 0$ and all $a_{-i} \in A_{-i}$, and $v(a^*) > v(a)$ for all $a \in A \setminus \{a^*\}$.

Such a function v is called a *monotone potential* for a^* in $\mathbf{g}$.

By definition, if a^* is a monotone potential maximizer, then $a_i^* = 0$ (resp. $a_i^* = 1$) for players $i \in I^0 \setminus I^1$ (resp. $i \in I^1 \setminus I^0$), and hence, a^* is an undominanted Nash equilibrium.

---

[3]Oyama and Takahashi (2020) referred to this version simply as monotone potential maximizer (without the qualifier "strict").

[4]See Proposition 3 and Lemma 9 in Morris and Ui (2005).

Clearly, a strict monotone potential maximizer is a monotone potential maximizer. The converse holds in generic games, but not in general: for example, in a game with constant payoffs where $f_i \equiv 0$ for all $i \in I$, all action profiles are trivially monotone potential maximizers (since $I^1 = I^0 = \emptyset$), while none of them is a strict monotone potential maximizer.

For the extreme action profiles, a monotone potential maximizer is a strict monotone potential maximizer in a restricted game. The following lemma states this observation for the smallest action profile $\mathbf{0}$, where $\mathbf{g}_{I^1}(\cdot, \mathbf{0}_{I \setminus I^1})$ denotes the restricted game among the players in I^1 where the action of every player in $I \setminus I^1$ is fixed to 0, so that the payoff increment function of player $i \in I^1$ is given by $f_i(\cdot, \mathbf{0}_{I \setminus I^1})$.

Lemma 1. *For a binary-action game* $\mathbf{g}$*,* $\mathbf{0}$ *is a monotone potential maximizer in* $\mathbf{g}$ *if and only if* $\mathbf{0}_{I^1}$ *is a strict monotone potential maximizer in* $\mathbf{g}_{I^1}(\cdot, \mathbf{0}_{I \setminus I^1})$*.*

By convention, the latter condition is vacuously true if $I^1 = \emptyset$, in which case $\mathbf{0}$ is trivially a monotone potential maximizer.

Proof. The "only if" direction is immediate. To prove the "if" direction, suppose that $I^1 \neq \emptyset$ and that $v^1 \colon A_{I^1} \to \mathbb{R}$ is a strict monotone potential for $\mathbf{0}_{I^1}$ in $\mathbf{g}_{I^1}(\cdot, \mathbf{0}_{I \setminus I^1})$: i.e., $v^1(\mathbf{0}_{I^1}) > v^1(b)$ for all $b \in A_{I^1} \setminus \{\mathbf{0}_{I^1}\}$, and there exists $(\lambda_i)_{i \in I^1}$ with $\lambda_i > 0$ for all $i \in I^1$ such that

$$\lambda_i f_i(b_{-i}, \mathbf{0}_{I \setminus I^1}) \leq v^1(1, b_{-i}) - v^1(0, b_{-i})$$

for all $i \in I^1$ and all $b_{-i} \in A_{I^1 \setminus \{i\}}$. Define $v \colon A \to \mathbb{R}$ by

$$v(a) = \begin{cases} v^1(a_{I^1}) & \text{if } a_{I \setminus I^1} = \mathbf{0}_{I \setminus I^1}, \\ -\sum_{i \in I^1 : a_i = 0} \lambda_i \max_{a'_{-i} \in A_{-i}} f_i(a'_{-i}) - M & \text{otherwise} \end{cases}$$

for $a = (a_{I^1}, a_{I \setminus I^1}) \in A_{I^1} \times A_{I \setminus I^1}$, where M is a constant such that $v(\mathbf{0}) > v(a)$ for all $a \in A \setminus \{\mathbf{0}\}$. We claim that this function is a monotone potential for $\mathbf{0}$ in $\mathbf{g}$. Indeed, for any $i \in I^1$ and any $a_{-i} = (a_{I^1 \setminus \{i\}}, a_{I \setminus I^1}) \in A_{I^1 \setminus \{i\}} \times A_{I \setminus I^1}$, if $a_{I \setminus I^1} = \mathbf{0}_{I \setminus I^1}$, then

$$v(1, a_{-i}) - v(0, a_{-i}) = v^1(1, a_{I^1 \setminus \{i\}}) - v^1(0, a_{I^1 \setminus \{i\}}) \geq \lambda_i f_i(a_{I^1 \setminus \{i\}}, \mathbf{0}_{I \setminus I^1}) = \lambda_i f_i(a_{-i}),$$

while if $a_{I \setminus I^1} \neq \mathbf{0}_{I \setminus I^1}$, then

$$v(1, a_{-i}) - v(0, a_{-i}) = \lambda_i \max_{a'_{-i} \in A_{-i}} f_i(a'_{-i}) \geq \lambda_i f_i(a_{-i}),$$

as desired. □

2.3. **Sequential Obedience.** In this section, we describe the dual condition that characterizes nonexistence of a strict monotone potential for action profile $\mathbf{0}$ and derive the corresponding condition for monotone potential.[5]

Let Γ denote the set of all sequences of distinct players (including the null sequence $\emptyset$), and for each $i \in I$, let Γ_i denote the set of all sequences in Γ where player i is listed. For each $i \in I$ and $\gamma \in \Gamma_i$, let $a_{-i}(\gamma)$ denote 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. For a binary-action game $\mathbf{g}$, a distribution $\rho \in \Delta(\Gamma)$ satisfies *sequential obedience* in $\mathbf{g}$ if

$$\sum_{\gamma \in \Gamma_i} \rho(\gamma) f_i(a_{-i}(\gamma)) \geq 0 \tag{1}$$

for all $i \in I$.

Note that inequality (1) is obviously satisfied for players $i \in I$ such that $\rho(\Gamma_i) = 0$.

As shown in Oyama and Takahashi (2020, Lemma 2), by duality, $\mathbf{0}$ is not a strict monotone potential maximizer if and only if there exists $\rho \in \Delta(\Gamma \setminus \{\emptyset\})$ that satisfies sequential obedience, and under the supermodularity of $\mathbf{g}$, such a ρ exists that assigns positive probability only to permutations of players in some subset $S^* \subset I$, $S^* \neq \emptyset$. Combining that result with Lemma 1 in the previous section, we have a dual characterization of nonexistence of a monotone potential for $\mathbf{0}$. For $S \subset I$, let $\Gamma(S) \subset \Gamma$ denote the set of all sequences of distinct players in S and $\Pi(S) \subset \Gamma(S)$ denote the set of all permutations of players in S (thus $\Gamma(S) = \bigcup_{S' \subset S} \Pi(S')$).

Lemma 2. *For a binary-action game* $\mathbf{g}$*,*

(1) *either* $\mathbf{0}$ *is a monotone potential maximizer in* $\mathbf{g}$*, or* $I^1 \neq \emptyset$ *and there exists* $\rho \in \Delta(\Gamma(I^1) \setminus \{\emptyset\})$ *that satisfies sequential obedience in* $\mathbf{g}$*, but not both; and*

(2) *in the latter case, if* $\mathbf{g}$ *is supermodular, then there exists* $\rho \in \Delta(\Pi(S^*))$ *for some* $S^* \subset I^1$*,* $S^* \neq \emptyset$*, that satisfies sequential obedience in* $\mathbf{g}$*.*

3. A Characterization of Robustness

We now have a characterization of the robustness of extreme action profiles in all binary-action supermodular games. It is stated for the action profile $\mathbf{0}$, but it applies also to $\mathbf{1}$ by reversing the labels 0 and 1.

[5] Oyama and Takahashi (2020) stated the condition for action profile $\mathbf{1}$. Of course, the difference is only expositional; one is equivalently translated to the other by reversing the action labels 0 and 1.

Theorem 1. *For any binary-action supermodular game* $\mathbf{g}$, $\mathbf{0}$ *is robust in* $\mathbf{g}$ *if and only if it is a monotone potential maximizer in* $\mathbf{g}$.

Proof. The "if" part follows from Morris and Ui (2005). To prove the "only if" part, assume that $\mathbf{0}$ is not a monotone potential maximizer in $\mathbf{g}$. Then by Lemma 2, we can take $S^* \subset I^1$, $S^* \neq \emptyset$, and $\rho \in \Delta(\Pi(S^*))$ such that ρ satisfies sequential obedience in $\mathbf{g}$.

Fix any $\varepsilon \in (0, 1]$. By the sequential obedience of ρ, we have

$$\left(1 - \frac{\varepsilon}{2}\right) \sum_{\gamma \in \Gamma_i} \rho(\gamma) f_i(a_{-i}(\gamma)) + \frac{\varepsilon}{2|S^*|} f_i(\mathbf{1}_{-i}) > 0 \tag{2}$$

for all $i \in S^*$, where by supermodularity, $f_i(\mathbf{1}_{-i}) > 0$ for all $i \in S^* \subset I^1$. Let $\eta > 0$ be sufficiently small that $(1 - \varepsilon/2)(1 - \eta)^{|S^*|-1} \geq 1 - \varepsilon$, and

$$\left(1 - \frac{\varepsilon}{2}\right) \sum_{\gamma \in \Gamma_i} (1 - \eta)^{-\ell(i,\gamma)} \rho(\gamma) f_i(a_{-i}(\gamma)) + \frac{\varepsilon}{2|S^*|} (1 - \eta)^{-|S^*|} f_i(\mathbf{1}_{-i}) > 0 \tag{3}$$

for all $i \in S^*$, where

$$\ell(i, \gamma) = \begin{cases} \ell & \text{if there exists } \ell \in \{1, \dots, k\} \text{ such that } i_\ell = i, \\ \infty & \text{otherwise} \end{cases}$$

for $i \in I$ and $\gamma = (i_1, \dots, i_k) \in \Gamma$.

We construct the elaboration $(T, \tilde{P}, \mathbf{u})$ of $\mathbf{g}$ as follows. For each $i \in I$, let

$$T_i = \begin{cases} \{1, 2, \dots\} & \text{if } i \in S^*, \\ \{1, \infty\} & \text{otherwise.} \end{cases}$$

Define $P, Q \in \Delta(T)$ by

$$P(t) = \begin{cases} \eta(1 - \eta)^m \rho(\gamma) & \text{if there exist } m \in \mathbb{N} \text{ and } \gamma \in \Pi(S^*) \\ & \text{such that } t_i = m + \ell(i, \gamma) \text{ for all } i \in I, \\ 0 & \text{otherwise} \end{cases}$$

and

$$Q(t) = \begin{cases} \eta(1 - \eta)^{t_i - |S^*|}/|S^*| & \text{if there exists } i \in S^* \text{ such that} \\ & t_i \geq |S^*| \text{ and } t_j = 1 \text{ for all } j \neq i, \\ 0 & \text{otherwise} \end{cases}$$

for each $t = (t_i)_{i \in I} \in T$ (where $\mathbb{N} = \{0, 1, 2, \dots\}$), and let $\tilde{P} \in \Delta(T)$ be given by

$$\tilde{P}(t) = \left(1 - \frac{\varepsilon}{2}\right) P(t) + \frac{\varepsilon}{2} Q(t)$$

for each $t \in T$. Let $\mathbf{u}$ be such that players $i \in S^*$ of types $t_i \geq |S^*|$ and players $i \in I \setminus S^*$ of type $t_i = \infty$ know that their own payoffs are given by those of $\mathbf{g}$, and action 1 is a strictly dominant action for players $i \in S^*$ of types $t_i \leq |S^*| - 1$ and players $i \in I \setminus S^*$

of type $t_i = 1$. Observe that the event where $t_i \geq |S^*|$ for all $i \in S^*$ and $t_i = \infty$ for all $i \in I \setminus S^*$ has probability

$$\left(1 - \frac{\varepsilon}{2}\right)(1-\eta)^{|S^*|-1} \geq 1 - \varepsilon,$$

and therefore, this game is an ε-elaboration of $\mathbf{g}$. For $i \in S^*$, $\tau \geq |S^*|$, and $t_{-i} \in T_{-i}$, we write

$$S^\tau_{-i}(t_{-i}) = \{j \in S^* \setminus \{i\} \mid t_j \leq \tau - 1\} \cup \{j \in I \setminus S^* \mid t_j = 1\}.$$

Claim 1. *For any $i \in S^*$, $\tau \geq |S^*|$, and $S \subset I \setminus \{i\}$,*

$$\begin{aligned}
&\tilde{P}(S^\tau_{-i}(t_{-i}) = S | t_i = \tau)\\
&= \left[\begin{array}{l} \left(1-\frac{\varepsilon}{2}\right)(1-\eta)^{-|S|-1}\rho(\{\gamma \in \Pi(S^*) \mid a_{-i}(\gamma) = \mathbf{1}_S\}) \times \mathbb{I}_{S \subset S^* \setminus \{i\}} \\ + \frac{\varepsilon}{2|S^*|}(1-\eta)^{-|S^*|} \times \mathbb{I}_{S = I \setminus \{i\}} \end{array}\right] / C_i,
\end{aligned}$$

where $C_i = \left(1 - \frac{\varepsilon}{2}\right)\sum_{\ell=1}^{|S^*|}(1-\eta)^{-\ell}\rho(\{\gamma \in \Pi(S^*) \mid \ell(i,\gamma) = \ell\}) + \frac{\varepsilon}{2|S^*|}(1-\eta)^{-|S^*|} > 0$, and $\mathbb{I}_\varphi = 1$ if the statement φ is true and $\mathbb{I}_\varphi = 0$ otherwise.

Proof. We have

$$\begin{aligned}
\tilde{P}(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S) = &\left(1-\frac{\varepsilon}{2}\right)P(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S)\\
&+ \frac{\varepsilon}{2}Q(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S)
\end{aligned}$$

with

$$\begin{aligned}
&P(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S)\\
&\quad= \begin{cases} \eta(1-\eta)^{\tau-|S|-1}\rho(\{\gamma \in \Pi(S^*) \mid a_{-i}(\gamma) = \mathbf{1}_S\}) & \text{if } S \subset S^* \setminus \{i\},\\ 0 & \text{otherwise,}\end{cases}\\
&Q(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S)\\
&\quad= \begin{cases} \eta(1-\eta)^{\tau-|S|-1}\eta(1-\eta)^{\tau-|S^*|}/|S^*| & \text{if } S = I \setminus \{i\},\\ 0 & \text{otherwise,}\end{cases}
\end{aligned}$$

while we have

$$\begin{aligned}
\tilde{P}(t_i = \tau) = &\left(1-\frac{\varepsilon}{2}\right)\sum_{\ell=1}^{|S^*|}\eta(1-\eta)^{\tau-\ell}\rho(\{\gamma \in \Pi(S^*) \mid \ell(i,\gamma) = \ell\})\\
&+ \frac{\varepsilon}{2}\eta(1-\eta)^{\tau-|S^*|}/|S^*|.
\end{aligned}$$

Then arranging terms in $\tilde{P}(t_i = \tau,\ S^\tau_{-i}(t_{-i}) = S)/\tilde{P}(t_i = \tau)$, we have the expression as claimed. ∎

In $(T, \tilde{P}, \mathbf{u})$, action 1 is uniquely rationalizable of all players in S^* of any type, and hence, $\mathbf{0}$ is never played in any Bayesian Nash equilibrium. Indeed, first, action 1 is a strictly dominant action for players $i \in S^*$ of types $t_i \leq |S^*| - 1$ and players $i \in I \setminus S^*$ of type $t_i = 1$ by construction. For $\tau \geq |S^*|$, suppose that action 1 is uniquely rationalizable for all players in S^* of types $t_i \leq \tau - 1$. Then the expected payoff increment for a player $i \in S^*$ of type $t_i = \tau$ from action 1 is no smaller than

$$\sum_{S \subset I \setminus \{i\}} \tilde{P}(S^\tau_{-i}(t_{-i}) = S | t_i = \tau) f_i(\mathbf{1}_S)$$
$$= \left[\left(1 - \frac{\varepsilon}{2}\right) \sum_{\gamma \in \Gamma_i} (1-\eta)^{-\ell(i,\gamma)} \rho(\gamma) f_i(a_{-i}(\gamma)) + \frac{\varepsilon}{2|S^*|} (1-\eta)^{-|S^*|} f_i(\mathbf{1}_{-i}) \right] / C_i > 0,$$

where the equality follows from Claim 1 and the inequality from condition (3). Therefore, action 1 is uniquely rationalizable for $t_i = \tau$. Thus, by induction, action 1 is uniquely rationalizable of all players in S^* of all types. This shows that $\mathbf{0}$ is not robust in $\mathbf{g}$. □

In the proof above and the proof of Oyama and Takahashi (2020, Theorem 2), we need strict incentives in constructing desired elaborations to prevent the target action profile (which is $\mathbf{0}$ here) from being played with high probability. In Oyama and Takahashi (2020), we appealed to genericity and assumed a strict version of inequality (1) (in the current notation). Here, we instead leverage the property that the nonexistence of a monotone potential for $\mathbf{0}$ implies the existence of ρ that satisfies (1) and assigns positive probability only on sequences of players for whom $f_i(\mathbf{1}_{-i}) > 0$. We thus obtain the strict inequality (2), which turns out to be sufficient to establish the nonrobustness of $\mathbf{0}$.

In fact, Oyama and Takahashi (2020) assumed the existence of a distribution $\rho \in \Delta(\Gamma \setminus \{\emptyset\})$ that satisfies the following strengthening of (1):

$$\sum_{\gamma \in \Gamma_i} (1-\eta)^{-\ell(i,\gamma)} \rho(\gamma) f_i(a_{-i}(\gamma)) > 0 \tag{4}$$

for all $i \in I$ such that $\rho(\Gamma_i) > 0$ (where $\ell(i,\gamma)$ denotes the rank of player i in sequence γ as in the proof of Theorem 1 above). This condition is stronger than nonrobustness of $\mathbf{0}$ in general, as the following example shows:

Example 1. Consider the two-player binary-action supermodular game:

	0	1
0	0, 0	0, 0
1	0, 0	1, 0

.

In this game, action profile $\mathbf{0}$ is not robust, since for player 1, action 0 is a weakly dominated action. On the other hand, no distribution $\rho \in \Delta(\Gamma \setminus \{\emptyset\})$ satisfies inequality (4) for all $i \in I$ with $\rho(\Gamma_i) > 0$.

The equivalence in Theorem 1 does not extend to nonextreme action profiles, as the following example shows:

Example 2. Consider the three-player binary-action supermodular game from Oyama and Takahashi (2019, Section A.6):

	0	1
0	$2,2,0$	$0,0,0$
1	$0,0,0$	$0,0,0$

0

	0	1
0	$1,1,0$	$0,0,0$
1	$0,0,0$	$1,1,0$

1.

In this game, $I^1 = I^0 = \{1, 2\}$, and action profile $\mathbf{0}$ is a monotone potential maximizer with a monotone potential

	0	1
0	2	0
1	0	0

0

	0	1
0	1	0
1	0	1

1.

Therefore, $\mathbf{0}$ is robust in this game by Morris and Ui (2005). Oyama and Takahashi (2019, Proposition A.2) show that action profile $(0, 0, 1)$ is also robust.[6] On the other hand, $(0, 0, 1)$ is not a monotone potential maximizer.[7,8]

4. Alternative Versions of Robustness

In this section, we consider two alternative versions of robustness, approximate robustness (Haimanko and Kajii (2016)) and strict robustness (Morris et al. (2023)), and present their characterizations that apply to all (whether extreme or nonextreme) action profiles in all binary-action supermodular games.

[6] In fact, the argument in the proof there shows that any action distribution that assigns probability 1 to $(a_1, a_2) = (0, 0)$ is robust.

[7] For, if $(0, 0, 1)$ is a monotone potential maximizer with a monotone potential $v\colon A \to \mathbb{R}$ and $\lambda_1, \lambda_2 > 0$, then we have $-\lambda_1 \leq v(1, 0, 1) - v(0, 0, 1)$, $\lambda_1 \leq v(1, 1, 1) - v(0, 1, 1)$, $-\lambda_2 \leq v(0, 1, 1) - v(0, 0, 1)$, and $\lambda_2 \leq v(1, 1, 1) - v(1, 0, 1)$; these lead to $v(0, 0, 1) \leq v(1, 1, 1)$, which contradicts $v(0, 0, 1) > v(a)$ for all $a \neq (0, 0, 1)$.

[8] This action profile becomes an extreme action profile if we reverse the lables of player 3's actions, but then the game is no longer supermodular, so that Theorem 1 does not apply.

4.1. **Strict Robustness.** Strict robustness strengthens robustness of Kajii and Morris (1997) by allowing for a larger class of incomplete information perturbations, where with high probability, the players believe that their payoff functions are close to those in $\mathbf{g}$. Given an elaboration $(T, P, \mathbf{u})$ of $\mathbf{g}$, for $i \in I$ and $\eta \geq 0$, let $T_i^{g_i,\eta}$ denote the set of player i's types for which the payoffs differ from g_i by at most η in expectation:

$$T_i^{g_i,\eta} = \left\{ t_i \in T_i \;\middle|\; \sum_{t_{-i}\in T_{-i}} P(t_{-i}|t_i) \max_{a\in A} |u_i(a,(t_i,t_{-i})) - g_i(a)| \leq \eta \right\},$$

and write $T^{\mathbf{g},\eta} = \prod_{i\in I} T_i^{g_i,\eta}$. Elaboration $(T, P, \mathbf{u})$ is called an (ε, η)-*elaboration* of $\mathbf{g}$ if $P(T^{\mathbf{g},\eta}) \geq 1 - \varepsilon$. An ε-elaboration in the sense of Kajii and Morris (1997) is an $(\varepsilon, 0)$-elaboration in this sense.

Definition 5. An action profile $a^* \in A$ is *strictly robust* in $\mathbf{g}$ if for every $\delta > 0$, there exist $\varepsilon > 0$ and $\eta > 0$ such that every (ε, η)-elaboration $(T, P, \mathbf{u})$ of $\mathbf{g}$ has a Bayesian Nash equilibrium σ such that $\sum_{t\in T} P(t) \prod_{i\in I} \sigma_i(a_i^*|t_i) \geq 1 - \delta$.

By definition, if a^* is strictly robust, then it is robust.

Morris et al. (2023) provided a characterization of strict robustness for binary-action supermodular games in terms of strict monotone potential maximizer.

Theorem 2 (Morris et al. (2023, Theorem 2))**.** *For any binary-action supermodular game* $\mathbf{g}$ *and any action profile* $a^* \in A$*,* a^* *is strictly robust in* $\mathbf{g}$ *if and only if it is a strict monotone potential maximizer in* $\mathbf{g}$.

The "if" part of this theorem in fact holds for many-action supermodular games (Morris et al. (2023, Theorem 1)). The "only if" part is obtained through the tight connection between strict robustness and full implementation by information design due to Morris et al. (2024); see Morris et al. (2023, Appendices A.2–A.3) for details.

4.2. **Approximate Robustness.** Approximate robustness weakens robustness of Kajii and Morris (1997) by allowing for approximate equilibria in incomplete information perturbations.

Definition 6. An action profile $a^* \in A$ is *approximately robust* in $\mathbf{g}$ if for every $\eta > 0$ and $\delta > 0$, there exists $\varepsilon > 0$ such that every ε-elaboration $(T, P, \mathbf{u})$ of $\mathbf{g}$ has an η-Bayesian Nash equilibrium σ such that $\sum_{t\in T} P(t) \prod_{i\in I} \sigma_i(a_i^*|t_i) \geq 1 - \delta$.

By definition, if a^* is robust, then it is approximately robust.

We have the following characterization of approximate robustness for binary-action supermodular games, by combining results from Haimanko and Kajii (2016), Oyama and

Takahashi (2020), and Morris et al. (2023). For $\eta > 0$, denote by $B_\eta(\mathbf{g})$ the set of games $\mathbf{g}'$ such that $\max_{i\in I, a\in A} |g'_i(a) - g_i(a)| \leq \eta$.

Theorem 3. *For any binary-action supermodular game* $\mathbf{g}$ *and any action profile* $a^* \in A$*, the following conditions are equivalent:*

(1) a^* *is approximately robust in* $\mathbf{g}$.

(2) *For any* $\eta > 0$*, there exists* $\mathbf{g}' \in B_\eta(\mathbf{g})$ *such that* a^* *is robust in* $\mathbf{g}'$.

(3) *For any* $\eta > 0$*, there exists* $\mathbf{g}' \in B_\eta(\mathbf{g})$ *such that* a^* *is strictly robust in* $\mathbf{g}'$.

(4) *For any* $\eta > 0$*, there exists* $\mathbf{g}' \in B_\eta(\mathbf{g})$ *such that* a^* *is a strict monotone potential maximizer in* $\mathbf{g}'$.

The implications (3) $\Rightarrow$ (2) $\Rightarrow$ (1) hold for general games: (2) $\Rightarrow$ (1) is by Haimanko and Kajii (2016, Theorem 3), and (3) $\Rightarrow$ (2) is immediate from the definition. The implication (4) $\Rightarrow$ (3) is, as already stated in Section 4.1, by Morris et al. (2023, Theorem 1) and holds for many-action supermodular games.

To show the implication (1) $\Rightarrow$ (4), we need to introduce a few more definitions. A distribution $\rho \in \Delta(\Gamma)$ satisfies *strict sequential obedience* in a binary-action game $\mathbf{g}$ if it satisfies the strict version of condition (1) in Definition 4, i.e.,

$$\sum_{\gamma\in\Gamma_i} \rho(\gamma) f_i(a_{-i}(\gamma)) > 0 \tag{5}$$

for all $i \in I$ such that $\rho(\Gamma_i) > 0$; $\rho \in \Delta(\Gamma)$ satisfies *reverse sequential obedience* (resp. *strict reverse sequential obedience*) in $\mathbf{g}$ if

$$\sum_{\gamma\in\Gamma_i} \rho(\gamma) f_i(a^0_{-i}(\gamma)) \leq \text{(resp. <) } 0 \tag{6}$$

for all $i \in I$ such that $\rho(\Gamma_i) > 0$, where $a^0_{-i}(\gamma) \in A_{-i}$ denotes the action profile of player i's opponents such that player $j \neq i$ plays action 0 if and only if j is listed in γ before i. Fix $a^* \in A$, and denote $S^* = \{i \in I \mid a^*_i = 1\}$. Let $\mathbf{g}_{S^*}(\cdot, \mathbf{0}_{I\setminus S^*})$ denote the restricted game of $\mathbf{g}$ among the players in S^* where the action of every player in $I \setminus S^*$ is fixed to 0. Symmetrically, let $\mathbf{g}_{I\setminus S^*}(\cdot, \mathbf{1}_{S^*})$ denote the restricted game among the players in $I \setminus S^*$ where the action of every player in S^* is fixed to 1.

Now, suppose that a^* is approximately robust in a binary-action supermodular game $\mathbf{g}$. First, the proof of Corollary A.1(2) in Oyama and Takahashi (2020) in fact shows (by contrapositive) that it implies that (i) there exists no $\rho \in \Delta(\Gamma(S^*) \setminus \{\emptyset\})$ that satisfies strict reverse sequential obedience in $\mathbf{g}_{S^*}(\cdot, \mathbf{0}_{I\setminus S^*})$ and (ii) there exists no $\rho \in \Delta(\Gamma(I \setminus S^*) \setminus \{\emptyset\})$ that satisfies strict sequential obedience in $\mathbf{g}_{I\setminus S^*}(\cdot, \mathbf{1}_{S^*})$. Next, for any

$\eta > 0$, let $\mathbf{g}' = (g'_i)_{i \in I} \in B_\eta(\mathbf{g})$ be the binary-action supermodular game defined by, for all $i \in I$ and all $a_{-i} \in A$,

$$g'_i(1, a_{-i}) = \begin{cases} g_i(1, a_{-i}) + \eta & \text{if } i \in S^*, \\ g_i(1, a_{-i}) - \eta & \text{if } i \in I \setminus S^*, \end{cases}$$

and $g'_i(0, a_{-i}) = g_i(0, a_{-i})$. Then it follows from conditions (i) and (ii) that (i′) there exists no $\rho \in \Delta(\Gamma(S^*) \setminus \{\emptyset\})$ that satisfies reverse sequential obedience in $\mathbf{g}'_{S^*}(\cdot, \mathbf{0}_{I \setminus S^*})$ and (ii′) there exists no $\rho \in \Delta(\Gamma(I \setminus S^*) \setminus \{\emptyset\})$ that satisfies sequential obedience in $\mathbf{g}'_{I \setminus S^*}(\cdot, \mathbf{1}_{S^*})$.[9] Finally, by Lemmas 2 and A.1 in Oyama and Takahashi (2020), conditions (i′) and (ii′) hold if and only if a^* is a strict monotone potential maximizer in $\mathbf{g}'$.

It is left as an open problem to determine whether Theorems 2 and 3 extend to many-action (supermodular) games.

References

Haimanko, O. and A. Kajii (2016): "Approximate Robustness of Equilibrium to Incomplete Information," *International Journal of Game Theory*, 45, 839–857.

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

——— (2020): "Notes on "Refinements and Higher Order Beliefs"," *Japanese Economic Review*, 71, 35–41.

Morris, S., D. Oyama, and S. Takahashi (2023): "Strict Robustness to Incomplete Information," *Japanese Economic Review*, 74, 357–376.

——— (2024): "Implementation via Information Design in Binary-Action Supermodular Games," *Econometrica*, 92, 775–813.

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

Oyama, D. and S. Takahashi (2019): "Generalized Belief Operator and the Impact of Small Probability Events on Higher Order Beliefs," SSRN 3375777.

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

[9] For, if (i′) fails, i.e., if some $\rho \in \Delta(\Gamma(S^*) \setminus \{\emptyset\})$ satisfies reverse sequential obedience in $\mathbf{g}'_{S^*}(\cdot, \mathbf{0}_{I \setminus S^*})$, then for all $i \in S^*$ such that $\rho(\Gamma_i) > 0$, we have $\sum_{\gamma \in \Gamma_i} \rho(\gamma)(g_i(1, a^0_{-i}(\gamma)) - g_i(0, a^0_{-i}(\gamma))) = \sum_{\gamma \in \Gamma_i} \rho(\gamma)(g'_i(1, a^0_{-i}(\gamma)) - g'_i(0, a^0_{-i}(\gamma))) - \rho(\Gamma_i)\eta \leq -\rho(\Gamma_i)\eta < 0$, so that ρ satisfies strict reverse sequential obedience in $\mathbf{g}_{S^*}(\cdot, \mathbf{0}_{I \setminus S^*})$, which contradicts (i); a symmetric argument applies to (ii′).

Oyama, D., S. Takahashi, and J. Hofbauer (2008): "Monotone Methods for Equilibrium Selection under Perfect Foresight Dynamics," *Theoretical Economics*, 3, 155–192.