Mixed bayesian implementation in general environments

Takashi KUNIMOTO
Singapore Management University, tkunimoto@smu.edu.sg

Published in Journal of Mathematical Economics,
Vol 82, May 2019, Pages 247-263
DOI 10.1016/j.jmateco.2019.03.003

Mixed Bayesian implementation in general environments☆

Takashi Kunimoto

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 5 October 2014
Received in revised form 14 March 2019
Accepted 15 March 2019
Available online 26 March 2019

Keywords:
Bayesian implementation
Closure
Incentive compatibility
Mixed Bayesian monotonicity
Nash implementation
Social choice set

ABSTRACT

A social choice rule is said to be mixed Bayesian implementable if one can design a mechanism (or institution) in which the set of all mixed Bayesian Nash equilibrium outcomes coincides with that specified by the rule. The objective of this paper is to generalize the results of mixed Bayesian implementation. By means of example, I first assess the implication of common priors in Bayesian implementation. Second, I identify a mild condition that fills the gap between the necessity and sufficiency for mixed Bayesian implementation in general environments including non-economic ones. Third, I establish some new results to unify the literature of Bayesian implementation and Nash implementation.

1. Introduction

The theory of implementation or mechanism design attempts to identify the conditions under which a social choice rule (or welfare criterion) may be decentralized through some institution (or mechanism). In contexts in which the planner knows what agents' preferences and/or beliefs (henceforth, I call them *types* collectively) might be, but does not know what they actually are, the theory has uncovered necessary and sufficient conditions for such decentralization.[1]

We say that a social choice rule is *partially implementable* by some mechanism if the mechanism possesses a Bayesian Nash equilibrium whose outcome is contained in that specified by the rule. We often appeal to the *revelation principle*, which says that whenever partial implementation is possible, one can always duplicate the same equilibrium outcome by using the *truthful* equilibrium in the *direct revelation* mechanism where each agent announces his type to the planner. Thus, a necessary condition for the implementation of any welfare criterion is its *incentive compatibility*: the best thing for each individual to do in the direct revelation mechanism is to report his true type as long as all other individuals truthfully announce their type.

Although the revelation principle has been a powerful tool in many applications, it is important to realize that the direct-revelation mechanism may possess other *untruthful* equilibria whose outcomes are not consistent with the welfare criterion. In order to take seriously the problems resulting from the multiplicity of equilibria, some researchers have turned to the question of *full Bayesian implementation*, and explored the conditions under which the *set* of Bayesian Nash equilibrium outcomes coincides with a given welfare criterion. In the case of full implementation, *Bayesian monotonicity* emerges, in addition to incentive compatibility. Indeed, full implementation is the concept of implementation this paper adopts.

The main objective of this paper is to generalize the results of full Bayesian implementation established in Jackson (1991) and Serrano and Vohra (henceforth, SV, 2010). Theorem 1 of Jackson (1991) shows that under the *economic condition*, which basically says that at least two agents can never be satiated, a social choice rule satisfies incentive compatibility, Bayesian monotonicity, and closure (to be defined in Section 5) if and only if it is fully *Bayesian implementable*. Jackson's Theorem 1 restricts attention to the

☆ This paper is formerly titled "Interim Equilibrium Implementation." This work was initiated while I was a visiting scholar at NUS (National University of Singapore). I am gratefully thankful to NUS Economics department for its hospitality. I am indebted to anonymous referees and the editor of the journal for detailed comments and suggestions that significantly improved the paper. This paper also has benefited greatly from the comments of Yi-Chun Chen, Michihiro Kandori, Hitoshi Matsushima, Roberto Serrano, Satoru Takahashi, Qianfeng Tang, Rajiv Vohra, Takuma Wakayama, and seminar participants at Academia Sinica, AMES in Singapore, Concordia University, Keio University, NUS, SAET Conference in Paris, SHUFE, and University of Tokyo. I acknowledge financial support from JSPS KAKENHI Grant Numbers 24330078 and 25780128, Japan Center for Economic Research, and Yamada Academic Research Fund.

E-mail address: tkunimoto@smu.edu.sg.

[1] For surveys on implementation theory, see, for example, Jackson (2001) and Serrano (2004).

setup where (1) each agent only uses pure strategies; (2) mechanisms are deterministic; and (3) the type space is finite. SV (2010) extend this result to the setup where the agents can use mixed strategies; mechanisms are stochastic; and the type space is quite general. More specifically, Theorem 1 of SV (2010) shows that in economic environments, mixed Bayesian implementation is equivalent to incentive compatibility, closure, and *mixed* Bayesian monotonicity, which is a strengthening of Bayesian monotonicity. The main contribution of this paper is to further generalize Theorem 1 of SV (2010) by dropping the economic condition. I consider this as a major addition to the literature because outside of economic environments, tight characterizations are generally not available, even for pure strategy equilibria. Note that Theorem 2 of Jackson (1991) proposes a sufficient condition for (pure) Bayesian implementation in "non-economic" environments. The table below depicts where the contribution of this paper lies in the literature:

	Pure strategies deterministic mechanisms finite type space	Mixed strategies stochastic mechanisms general type space
Economic environments	Theorem 1 of Jackson (1991)	SV (2010)
Non-Economic environments	Theorem 2 of Jackson (1991)	This paper

The main result of this paper is to identify a mild condition under which one can fully characterize mixed Bayesian implementation in general environments including non-economic ones. More formally:

Theorem. *Suppose that there are at least three agents and a given social choice rule satisfies the no-worst-rule condition (NWR). Then, the social choice rule is mixed Bayesian implementable if and only if it satisfies incentive compatibility, mixed Bayesian monotonicity, and closure.*

The no-worst-rule condition (NWR) is considered a version of *no-total-indifference* condition, which basically says that the environment is rich enough so that every type never be indifferent over the outcomes. I will later argue that NWR is a mild condition and illustrate its permissiveness by means of an example (Example 2). Another aspect of the contribution of this paper is to propose a unified treatment among different setups: I handle (i) the case of two agents (Theorem 2), (ii) the case of "single-valued" social choice rules (Theorem 3), and (iii) the case of *complete information*, which describes the situation in which the underlying state is always commonly known among the agents. In complete information environments, I identify a condition much weaker than NWR under which one can fully characterize *mixed Nash implementation* in general environments including non-economic ones (Theorem 4).[2]

I move on to the last aspect of the generalization this paper executes. In most of applications in mechanism design and implementation theory, the researchers invoke the *common prior* assumption, which requires that all agents share the common belief about the state at the ex ante stage. I therefore investigate the implication of common priors in mixed Bayesian implementation. By Example 1, I confirm that if the common prior assumption is violated, Bayesian Nash equilibrium loses its predictive power quite a lot. In Section 6, building upon an example of Palfrey and Srivastava (1989b), I argue that mixed Bayesian monotonicity can be more permissive when the common prior assumption is violated than when it is satisfied.

The rest of the paper is organized as follows. Section 2 clarifies the scope of the current paper in the literature. In Section 3, I introduce the general setup for the paper. In Section 4, I introduce the concept of mixed Bayesian implementation. In Section 5, I identify the necessary conditions for implementation. In Section 6, I illustrate all the necessary conditions for implementation by means of an example. Section 7 provides a set of sufficient conditions for mixed Bayesian implementation. In Section 8, I restrict attention to complete information environments and obtain the sufficiency result for Nash implementation. Section 9 concludes. In the Appendix, I provide all the proofs omitted from the main body of the paper and discuss how one can extend this paper's analysis to a more general setup.

2. Related literature and the scope of the paper

The current paper contributes to unifying the literature of Bayesian implementation and Nash implementation. In what follows, I will be clear about all the aspects of the generalizations the current paper executes. However, those who are mainly interested in the main results of this paper can skip this section and immediately move to Section 3.

2.1. Type space

Assuming that the payoff-relevant parameter space is Polish, Brandenburger and Dekel (1993) construct the so-called *universal type space* that consists of all coherent belief hierarchies, which is also Polish.[3] Since this paper's type space can be Polish, it becomes quite general, as it can be interpreted as the universal type space. Duggan (1997) and SV (2010) consider an even more general type space. This paper's topological assumption is only needed for the sufficiency results in Sections 7 and 8. In contrast, Jackson (1991) restricts attention to a finite type space.

This paper assumes that the underlying type space is common knowledge among the planner and the agents. This common knowledge assumption is often seen as unrealistic so that one seeks for *robust implementation*, which requires that implementation survive any specification of higher-order beliefs consistent with the common knowledge structure of the environment. See Artemov et al. (2013) and Bergemann and Morris (2009, 2011) for robust implementation.

2.2. Multi-valued social choice rules

Many researchers focus on single-valued social choice rules, i.e., social choice *functions* (SCFs). However, the restriction to functions often turns out to be severe constraints on its implementability (See Muller and Satterthwaite (1977) and Saijo (1987)). Therefore, to obtain more permissive implementation results, the literature considers multi-valued social choice rules, i.e., social choice *sets* (SCSs). In fact, many interesting SCSs satisfy Maskin monotonicity: the Pareto, Core, Walrasian, Envy-Free, Lindhal, all these SCSs satisfy it. On the other hand, any social choice "function" selected from these SCSs no longer satisfies Maskin monotonicity. Moreover, Bayesian monotonicity is known

[2] See Section 8 for the precise definition of Nash implementation.

[3] A Polish space is a separable, completely metrizable topological space.

to be a quite restrictive condition even for SCSs.[4] The current paper handles the case of SCSs as well as SCFs.

2.3. Domain restrictions

The property upon which this paper relies is the no-worst-rule condition (NWR). Note that NWR is stronger than the *no-total-indifference* condition used by SV (2010). While this paper does not need it at all, SV (2010) need Jackson's (1991) *economic condition*, which essentially says that at least two agents can never be simultaneously satiated. Moreover, Abreu and Matsushima (1992, 1994) use an assumption that plays essentially the same role of small side payments.[5] Theorems 1 and 2 of the current paper show that one can do away with the economic condition but rather need a version of no-total-indifference condition.

2.4. Mechanisms

While many classical papers only deal with "deterministic" mechanisms, some papers consider "stochastic" mechanisms.[6] This paper's stochastic mechanisms draw upon the constructions proposed by Bergemann et al. (2011) and Oury and Tercieux (2012).

The current paper exploits the nature of *infinite* mechanisms. One obvious benefit of utilizing infinite mechanisms is that one can close the gap between the necessity and sufficiency for implementability. See Abreu and Matsushima (1992, 1994) and Chen et al. (2016) for the use of finite implementing mechanisms.

2.5. Mixed strategy

The theory of implementation has often left out the consideration of mixed strategy equilibria. This is particularly problematic for a research program that attempts to address the problem of multiplicity of equilibria in mechanisms.[7] Jackson (1992, Section 5.2) argues most eloquently that ignoring mixed strategies becomes a serious problem.

SV (2010) seriously tackle the issue of mixed strategy equilibria by stochastic mechanisms. On the other hand, Maskin (1999) uses deterministic mechanisms, but requires each outcome in the support of any mixed strategy equilibrium to be in the social choice rule. This paper follows the approach of SV (2010). In virtual implementation, Abreu and Matsushima (1992) handle mixed strategies explicitly and Duggan (1997) argues that the difference between pure and mixed strategies goes away because one can construct a sufficiently large type space and perform a purification of any mixed strategy equilibrium.[8] In robust implementation, Bergemann and Morris (2009, 2011) and Artemov et al. (2013) appeal to a similar argument of Duggan (1997).

2.6. Complete information or not

Historically, the literature of implementation theory pays special attention to *complete information* environments in which the underlying state is always commonly certain among all the agents. The current paper deals with incomplete information environments and treats complete information ones as a special case of the former. From this unified perspective, I show that moving from incomplete to complete information environments, one can weaken mixed Bayesian monotonicity and NWR into Maskin monotonicity and the *no-worst-alternative (NWA)* condition (to be defined in Section 8), respectively.

3. Preliminaries

I consider the implementation problem in environments with incomplete information. Let $N = \{1, \dots, n\}$ denote the finite set of agents and T_i be the set of types of agent i. Throughout the paper, I assume that T_i is a finite space.[9] Denote $T \equiv T_1 \times \cdots \times T_n$, and $T_{-i} \equiv T_1 \times \cdots \times T_{i-1} \times T_{i+1} \times \cdots \times T_n$.[10] Let $\Delta(T_{-i})$ denote the set of probability distributions on T_{-i}. Each agent i has a system of "interim" beliefs that is expressed as a function $\pi_i : T_i \to \Delta(T_{-i})$. Then, I call $(T_i, \pi_i)_{i\in N}$ a *type space*. Let A denote a finite set of pure outcomes, which are assumed to be independent of the information state.[11] Let $\Delta(A)$ be the set of probability distributions on A. Agent i's state dependent von Neumann–Morgenstern utility function is denoted $u_i : \Delta(A) \times T \to \mathbb{R}$. I now define an *environment* as $\mathcal{E} = (A, \{u_i, T_i, \pi_i\}_{i\in N})$, which is implicitly understood to be common knowledge among the agents.

A subset of T is called an event. An event $E = E_1 \times \cdots \times E_n \subseteq T$ is said to be *belief-closed* if, for each $i \in N$ and $t_i \in E_i$, we have $\sum_{t_{-i}\in E_{-i}} \pi_i(t_i)[t_{-i}] = 1$. In words, if an event E is a belief-closed subspace, it is commonly certain among all agents that E obtains. Throughout the paper I assume that the planner (or mechanism designer) only cares about the subset of the type space $(T_i^*)_{i\in N}$ where $T_i^* \subseteq T_i$. This paper takes $(T_i^*, \pi_i)_{i\in N}$ as an arbitrarily belief-closed subspace of $(T_i, \pi_i)_{i\in N}$.[12] As in the case of complete information settings, for example, T^* may be a proper subset of T.

A (stochastic) *social choice function* (SCF) is a function $f : T \to \Delta(A)$. Let $\mathbb{F} = \{f \mid f : T \to \Delta^*(A)\}$ be the set of SCFs, where $\Delta^*(A)$ is defined as a finite subset of $\Delta(A)$. The finiteness of $\mathbb{F}$ is imposed simply to avoid the measurability issue.[13] This paper is mainly concerned with *social choice sets* (SCSs), which are "multi-valued" social choice rules. An SCS F is defined as a nonempty subset of $\mathbb{F}$. Since $\mathbb{F}$ is finite, any SCS F, which is a nonempty subset of $\mathbb{F}$, is also finite. Two SCSs F and H are said to be *equivalent* ($F \approx H$) if there exists a bijection $\xi : F \to H$ such that for every $f \in F$ and every $h \in H$ satisfying $h = \xi(f)$, $f(t) = h(t)$ for all $t \in T^*$. This means that the two SCSs "coincide" for every $t \in T^*$.

A *mechanism* (or *game form*) $\Gamma = ((M_i)_{i\in N}, g)$ describes a nonempty countable message space M_i for each agent i and an

4 For instance, Chakravorti (1992) and Palfrey and Srivastava (1987) show that Bayesian monotonicity is very restrictive even for the case of SCSs.

5 Furthermore, Matsushima (1993) explicitly assumes that side payments are available to show that Bayesian monotonicity is generically a vacuous constraint and Artemov et al. (2013) and Bergemann and Morris (2009) also need a similar assumption of Abreu and Matsushima (1992) for robust virtual implementation.

6 For stochastic mechanisms, see also Matsushima (1993) and SV (2010) in Bayesian implementation; Bergemann et al. (2011) in implementation in rationalizable strategies; Bergemann and Morris (2011) in robust exact implementation; and Artemov et al. (2013) and Bergemann and Morris (2009) in robust virtual implementation.

7 In fact, Jackson (1991), Postlewaite and Schmeidler (1986), Mookherjee and Reichelstein (1990) and Palfrey and Srivastava (1987, 1989a) all consider "pure" strategies only.

8 A mixed strategy equilibrium is *purified* if there exist a sequence of Bayesian games and a sequence of pure strategy equilibria such that the mixed strategy equilibrium is close to the limit of the associated sequence of pure strategy equilibria.

9 In the Appendix, I discuss the extension of my analysis to a *Polish* space T_i associated with its Borel σ-algebra $\mathscr{T}_i$.

10 Similar notation will be used for products of other sets.

11 In the Appendix, I extend the analysis to a more general set of A associated with its σ-algebra $\mathscr{A}$ containing all singleton sets.

12 For example, Jackson (1991) assumes that all agents have a common support prior over T. Then, T^* is interpreted as the set of profiles of types to which agents assign strictly positive probability. SV (2010) also make the same assumption.

13 In the Appendix, I define $\mathbb{F} = \{f \mid f : T \to \Delta(A)\}$ as the set of all SCFs such that $\Delta^*(A)$ is replaced by $\Delta(A)$ and T is replaced by $(T_i, \mathscr{T}_i)_{i\in N}$.

outcome function $g : M \to \Delta(A)$, where $M = \times_{i\in N} M_i$.[14] Let $\Gamma^{DR} = ((T_i)_{i\in N}, f)$ denote the *direct revelation* mechanism associated with an SCF f.

The interim expected utility of agent i of type t_i that pretends to be of type t_i' in the direct-revelation mechanism associated with an SCF f, provided all other agents are truthful is defined as:

$$U_i(f; t_i'|t_i) \equiv \sum_{t_{-i}} \pi_i[t_i](t_{-i}) u_i(f(t_i', t_{-i}); (t_i, t_{-i}).$$

Denote $U_i(f|t_i) = U_i(f; t_i|t_i)$.

4. Mixed Bayesian implementation

I shall introduce the concept of mixed Bayesian implementation. Given a mechanism $\Gamma = (M, g)$, let $\Gamma(T)$ denote an incomplete information game associated with a type space $(T_i, \pi_i)_{i\in N}$. Let $\sigma_i(\cdot|t_i) \in \Delta(M_i)$ denote the probability measure over M_i conditional upon t_i. Besides, I denote by $\sigma(\cdot|t) = \prod_{j\in N} \sigma_j(\cdot|t_j) \in \prod_{j\in N} \Delta^*(M_j)$ the product measure over M conditional upon t. I denote by $\mathrm{supp}(\sigma(\cdot|t))$ the *support* of probability measure $\sigma(\cdot|t)$ on M. Similar notation will be used for other probability measures. I let

$$U_i(g \circ (\sigma_i', \sigma_{-i})|t_i) \equiv \sum_{t_{-i}} \pi_i[t_i](t_{-i}) \sum_{m_{-i}} \sigma_{-i}(t_{-i})[m_{-i}] \times \sum_{m_i} \sigma_i'(t_i)[m_i] u_i(g(m_i, m_{-i}); (t_i, t_{-i})).$$

Bayesian Nash equilibrium is a solution concept this paper adopts.

Definition 1 (*Bayesian Nash Equilibrium*)**.** A strategy profile $\sigma \in \Sigma$ is a **Bayesian Nash equilibrium** of the game $\Gamma(T)$ if, for each $i \in N$, $t_i \in T_i$, and strategy $\sigma_i' \in \Sigma_i$,

$$U_i(g \circ \sigma|t_i) \geq U_i(g \circ (\sigma_i', \sigma_{-i})|t_i).$$

Given an incomplete information game $\Gamma(T)$, let $BNE^{\Gamma(T)}$ be the collection of strategy profiles such that each $\sigma \in BNE^{\Gamma(T)}$ is a Bayesian Nash equilibrium of the game $\Gamma(T)$.

Here I introduce the concepts of common support priors and common priors as extra restrictions on agents' interim beliefs.

Definition 2. A type space $(T_i, \pi_i)_{i\in N}$ is said to be the one with **common support priors** if there exists a collection of priors $\{p_i\}_{i\in N}$ with each $p_i \in \Delta(T)$ satisfying the following two properties: (1) $\pi_i(t_i) = p_i(\cdot|t_i) \in \Delta(T_{-i})$ for any $i \in N$ and $t_i \in T_i$ where $p_i(\cdot|t_i)$ denotes the probability distribution on T_{-i} conditional upon t_i and (2) all agents' priors agree on the zero-probability events: for each $i, j \in N$ and each $\hat{t} \in T$, $p_i(\hat{t}) = 0$ if and only if $p_j(\hat{t}) = 0$. Moreover, a type space $(T_i, \pi_i)_{i\in N}$ is said to be the one with a **common prior** if it is the one with common support priors and there exists $p \in \Delta(T)$ such that $p_i(\cdot) = p(\cdot)$ for each $i \in N$.

Remark. The common support assumption means that at the ex ante stage, all agents agree on the zero-probability events. Lipman (2003) shows that this common support condition is equivalent to the common prior assumption in a "finite" type space.

I illustrate the role of common prior as well as common supports in Bayesian Nash equilibrium by means of an example. By this example, I do not claim its novelty. The whole purpose of this example is to make a simple point. That is, the set of BNE on a common prior space differs significantly from the set of BNE on a space without a common prior. I connect the role of common priors and common supports to the necessary conditions of mixed Bayesian implementation in Section 6.

Example 1. We consider an incomplete information game $\Gamma(T)$ in which (1) $N = \{1, 2\}$ as the set of agents; (2) $A = \{a_1, a_2\}$ as the set of pure alternatives; (3) $M_i = \{m_i, m_i'\}$ for each $i \in N$ as the set of actions available for agent i; (4) $g(m_1, m_2) = g(m_1', m_2') = a_1$ and $g(m_1, m_2') = g(m_1', m_2) = a_2$; (5) $T_i = \{t_i, t_i'\}$ as the set of types for each agent $i \in N$; (4) each agent's utility function is given: $u_1(a_1; t) = 1 > -1 = u_1(a_2; t)$ and $u_2(a_1; t) = -1 < 1 = u_2(a_2; t)$ for each $t \in T = T_1 \times T_2$; and (6) each agent i's interim beliefs over T_{-i} are: $\pi_1(t_1)[t_2] = 1$; $\pi_1(t_1')[t_2'] = 1$; $\pi_2(t_2)[t_1'] = 1$; and $\pi_2(t_2')[t_1] = 1$.

Note that the type space T is a belief-closed space in which the two agents have beliefs with completely disjoint support., i.e., the common support assumption is clearly violated. Consider the following strategy profile σ: $\sigma_1(t_1) = m_1$; $\sigma_1(t_1') = m_1'$; $\sigma_2(t_2) = m_2$; and $\sigma_2(t_2') = m_2'$. It is easy to see that σ constitutes a Bayesian Nash equilibrium because each σ_i achieves the best possible payoff for every type against σ_{-i} given his interim belief π_i.

Since the agents' utility functions are commonly certain between the agents, each type $t_i \in T_i$ plays the role of a payoff-irrelevant signal only observable to agent i. This means that Bayesian Nash equilibrium σ here can be considered a subjective correlated equilibrium in which the type space constitutes a correlating device. The underlying complete information game is a 2×2 zero-sum game so that each player has the unique optimal strategy where he chooses each pure action with equal probability. It is well known that in a two-person zero-sum game, all correlated equilibria are convex combinations of pairs of optimal strategies.[15] Therefore, in this example, there is the unique correlated equilibrium in which each player uses his unique optimal strategy. We therefore conclude that once we impose the common prior assumption on T, σ no longer constitutes a Bayesian Nash equilibrium.

When the set of Bayesian Nash equilibrium outcomes is required to exactly coincide with those picked by the SCS, I speak of *mixed Bayesian implementation*, which is proposed by SV (2010).

Definition 3 (*Mixed Bayesian Implementation*)**.** An SCS F is **mixed Bayesian implementable** if there exists a mechanism $\Gamma = (M, g)$ such that $g \circ BNE^{\Gamma(T)} \approx F$. More specifically, this requirement can be decomposed into the following two properties: (1) for each $f \in F$, there exists $\sigma \in BNE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for each $\sigma \in BNE^{\Gamma(T)}$, there exists $\hat{f} \in F$ such that $g \circ \sigma \approx \hat{f}$.

5. Necessity for mixed Bayesian implementation

In this section, I discuss the necessary conditions for mixed Bayesian implementation. First, I start from *incentive compatibility*.

Definition 4. An SCS F satisfies **incentive compatibility** if, for each $f \in F$, $i \in N$, and $t_i, t_i' \in T_i^*$,

$$U_i(f|t_i) \geq U_i(f; t_i'|t_i).$$

The proposition below is already proved by Jackson (1991).

Proposition 1. *If an SCS F is mixed Bayesian implementable, it satisfies incentive compatibility.*

[14] In the Appendix, I discuss how the analysis can be extended to a more general class of mechanisms.

[15] See, for example, footnote 14 of Aumann (1987).

A *deception* is a collection $\beta = (\beta_i)_{i\in N}$, where each $\beta_i : T_i \to 2^{T_i}\backslash\{\emptyset\}$. Let $\beta(t) = (\beta_1(t_1), \ldots, \beta_n(t_n))$. Write $f \circ \beta \in F$ if there exist $\tilde{F}$ and $\tilde{f} \in \tilde{F}$ such that $F \approx \tilde{F}$ and $f \circ \beta \approx \tilde{f}$. Otherwise, Write $f \circ \beta \notin F$. I introduce *mixed Bayesian monotonicity* proposed by SV (2010).

Definition 5. An SCS F satisfies **mixed Bayesian monotonicity** if, for every $f \in F$, every deception β for which $f \circ \beta \notin F$, and every collection of $\{\psi_k\}_{k\in N}$ with each $\psi_k : T_k \to \Delta(T_k)$ such that for each $k \in N$ and $t_k \in T_k$,

$$\psi_k(t'_k|t_k) > 0 \Leftrightarrow t'_k \in \beta_k(t_k),$$

there exist $i \in N,\ t_i \in T_i^*$, and a function $y^* : T_{-i} \to \Delta(A)$ such that

$$U_i(y^* \circ \psi_{-i}|t_i) > U_i(f \circ \psi|t_i)$$

where $\psi_{-i} = \prod_{j\neq i} \psi_j : T_{-i} \to \prod_{j\neq i} \Delta(T_k)$ and for all $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^*|\tilde{t}_i).$$

The direct-revelation mechanism might possess untruthful equilibria as well as the truthful one. Then, mixed Bayesian monotonicity can be considered a condition that knocks out untruthful equilibria without upsetting the truthful equilibrium. This implies that, with the help of mixed Bayesian monotonicity, we are guided clearly how to augment the direct revelation mechanism by adding "objection" as an extra message to the right agent (a whistle-blower) so that we sustain the truthful equilibrium, while upsetting untruthful equilibria in the augmented mechanism. This augmentation process is reflected in the construction of the mechanism in the proof of Theorem 1.

To fix the idea, imagine that each agent k employs a mixed strategy ψ_k in the direct revelation mechanism and β describes the support of the strategy profile ψ. Whenever ψ generates an undesirable outcome (i.e., $f\circ\beta \notin F$), mixed Bayesian monotonicity guarantees the existence of a test agent i of type t_i and an SCF y^* such that type t_i can be a whistle-blower by being provided y^* rather than $f \circ \psi$ if and only if the other agents in fact play ψ_{-i}. The next proposition for the necessity of mixed Bayesian monotonicity is already proved by SV (2010).

Proposition 2. *If an SCS F is mixed Bayesian implementable, then it satisfies mixed Bayesian monotonicity.*

I introduce a simpler class of deceptions which is often considered in the literature. β is called a *single-valued deception* if it is a deception and $\beta_i : T_i \to T_i$ for each $i \in N$. If only single-valued deceptions are considered in the definition of mixed Bayesian monotonicity, I call the corresponding concept simply Bayesian monotonicity used by Jackson (1991).

Definition 6. An SCS F satisfies **Bayesian monotonicity** if, for every $f \in F$ and every single-valued deception β for which $f\circ\beta \notin F$, there exist $i \in N,\ t_i \in T_i^*$, and a function $y^* : T_{-i} \to \Delta(A)$ such that

$$U_i(y^* \circ \beta_{-i}|t_i) > U_i(f \circ \beta|t_i),$$

where $\beta_{-i} = \prod_{j\neq i} \beta_j$ and for all $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^*|\tilde{t}_i).$$

SV (2010) introduce the following condition which clarifies the difference between mixed Bayesian monotonicity and Bayesian monotonicity.

Definition 7. An SCS F satisfies the **convex range property** if, whenever it is true that for $f \in F$ and a collection of single-valued deceptions $\{\beta^\lambda\}$ indexed by λ, one has that $f \circ \beta^\lambda \in F$ for each λ, then it is true that for every (not necessarily single-valued) deception $\tilde{\beta}$, if $\tilde{\beta} = \bigcup_\lambda \beta^\lambda$, then one has that $f \circ \tilde{\beta} \in F$.

Remark. It is easy to see that every "SCF" trivially satisfies this property.

I show an equivalence between mixed Bayesian monotonicity and Bayesian monotonicity.[16]

Proposition 3. *Suppose that an SCS F satisfies the convex range property and incentive compatibility. Then, F satisfies mixed Bayesian monotonicity if and only if it satisfies Bayesian monotonicity.*

Proof. The proof is in the Appendix. ■

I need some preparation for introducing the last necessary condition I consider for SCSs. For a belief-closed subspace $E \subseteq T$ and an SCS F, define

$$F(E) \equiv \left\{\alpha \in \Delta(A) \middle|\ \exists f \in F,\ \exists t \in E \text{ s.t. } f(t) = \alpha\right\}.$$

For two belief-closed subspaces $E, E' \subseteq T$ and an SCS F, define

$$F(E \times E') \equiv \left\{(\alpha, \alpha') \middle|\ \exists f \in F,\ \exists t \in E, \exists t' \in E' \text{ s.t. } f(t) = \alpha \text{ and } f(t') = \alpha'\right\}.$$

Definition 8. An SCS F satisfies **closure** if, for any pair of belief-closed subspaces $E, E' \subseteq T$, we have

$$F(E \times E') = F(E) \times F(E').$$

In words, closure says that Bayesian Nash equilibria should not depend upon any extra correlation between two belief-closed subspaces. The next result shows that closure is a necessary condition for mixed Bayesian implementation. See Jackson (1991) and Palfrey and Srivastava (1993) for the proof.

Proposition 4. *If an SCS F is mixed Bayesian implementable, it satisfies closure.*

6. An example

I build my argument on an important example (Example 3 of Palfrey and Srivastava, 1989a,b) to show the permissiveness of the necessary conditions for implementation.[17] There are two pure outcomes, $A = \{a, b\}$ and three agents, $N = \{1, 2, 3\}$. Each agent i has two possible types, $T_i = \{t_a, t_b\}$ and each type is drawn independently such that t_b is chosen with probability $q \in (0, 1)$ and I assume $q^2 > 1/2$. Therefore, for each $i \in N$ and $t_i \in T_i$, type t_i's interim belief $\pi_i[t_i]$ is provided as follows: for any $t_{-i} \in T_{-i}$,

$$\pi_i[t_i](t_{-i}) = \begin{cases} q^2 & \text{if } t_{-i} = (t_b, t_b) \\ q(1-q) & \text{if } t_{-i} = (t_b, t_a) \text{ or } (t_a, t_b) \\ (1-q)^2 & \text{if } t_{-i} = (t_a, t_a) \end{cases}$$

Note that these interim beliefs can be derived from a common prior $p \in \Delta(T)$ such that $p(t_b, t_b, t_b) = q^3$; $p(t_b, t_b, t_a) = p(t_b, t_a, t_b) = p(t_a, t_b, t_b) = q^2(1-q)$; $p(t_b, t_a, t_a) = p(t_a, t_b, t_a) = p(t_a, t_a, t_b) = q(1-q)^2$; and $p(t_a, t_a, t_a) = (1-q)^3$. Agents have identical preferences such that for each $i \in N$ and $t \in T$,

$$u_i(a; t) = \begin{cases} 1 & \text{if at least two agents are of type } t_a, \\ 0 & \text{otherwise;} \end{cases}$$

[16] Proposition 1 of SV (2010) shows the same result but without incentive compatibility. I found that the logic of their argument does not go through without incentive compatibility. Here I only claim that incentive compatibility is sufficient and essential for this result but I do not claim its necessity.

[17] This example was also extensively discussed in Kunimoto and Serrano (2011) and SV (2005).

$$u_i(b,t)=\begin{cases}1 & \text{if at least two agents are of type } t_b,\\ 0 & \text{otherwise.}\end{cases}$$

Let $\mathcal{E}=(A,\{u_i,T_i,\pi_i\}_{i\in N})$ denote the corresponding environment. For each agent $i\in N$, the corresponding interim utilities for the constant SCFs assigning outcomes a and b are:

$U_i(a|t_a)=1-q^2$; $U_i(b|t_a)=q^2$; $U_i(a|t_b)=(1-q)^2$; and $U_i(b|t_b)=1-(1-q)^2$.

Since $q^2>1/2$, this implies that $U_i(b|t_i)>U_i(a|t_i)$ for all $i\in N$ and $t_i\in T_i$.

Set $T^*=T$ as the set of type profiles with which the planner is concerned. Consider the "majoritarian" SCF, $f:T\to A$, which chooses a when at least two agents are of type t_a and b when at least two agents are of type t_b. This SCF f clearly satisfies incentive compatibility and closure trivially. However, f does not satisfy Bayesian monotonicity, hence it is not implementable. To see this consider the single-valued deception $\beta_i(t_i)=t_b$ for all $i\in N$ and $t_i\in T_i$. Note first that $f\circ\beta\not\approx f$. Since $f(\beta(t))=b$ for all $t\in T$ and $U_i(b|t_i)>U_i(a|t_i)$ for all $i\in N$ and $t_i\in T_i$, there do not exist $i\in N$, $t_i\in T_i^*$, and $y:T_{-i}\to\Delta(A)$ such that $U_i(y\circ\beta|t_i)>U_i(f\circ\beta|t_i)$.

I now propose a slight modification of the previous example. I keep the same set of players ($N=\{1,2,3\}$), the same set of pure outcomes ($A=\{a,b\}$), and the same set of types for agents 1 and 2 but add one extra type t_c to the set of types of agent 3 so that T_3 now becomes $\{t_a,t_b,t_c\}$. Hence, I set $\hat{T}_i=\{t_a,t_b\}$ for each $i\in\{1,2\}$ and $\hat{T}_3=\{t_a,t_b,t_c\}$. For each agent $i\in N$ and $t_i\in\hat{T}_i$, each type t_i's interim belief $\hat{\pi}_i[t_i]$ is provided as the same as before: for any $t_{-i}\in\hat{T}_{-i}$,

$$\hat{\pi}_i(t_i)[t_{-i}]=\begin{cases}q^2 & \text{if } t_{-i}=(t_b,t_b)\\ q(1-q) & \text{if } t_{-i}=(t_b,t_a)\text{ or }(t_a,t_b)\\ (1-q)^2 & \text{if } t_{-i}=(t_a,t_a)\end{cases}$$

This implies that every type of agents 1 and 2 never believes that agent 3 is of type t_c. Therefore, this (slightly) expanded type space $\hat{T}$ is the one without common support priors. Agents have identical preferences such that for each $i\in N$ and $t\in\hat{T}$,

$$\hat{u}_i(a;t)=\begin{cases}1 & \text{if at least two agents are of type } t_a \text{ and } t_3\neq t_c,\\ 1 & \text{if } t_3=t_c,\\ 0 & \text{otherwise;}\end{cases}$$

$$\hat{u}_i(b,t)=\begin{cases}1 & \text{if at least two agents are of type } t_b \text{ and } t_3\neq t_c,\\ 0 & \text{otherwise,}\end{cases}$$

Let $\hat{\mathcal{E}}=(A,\{\hat{u}_i,\hat{T}_i,\hat{\pi}_i\}_{i\in N})$ denote the corresponding environment. Set $\hat{T}=T^*$ as the set of type profiles with which the planner is concerned. I slightly modify the previous SCF f into $\hat{f}$ as follows: for any $t\in\hat{T}$,

$$\hat{f}(t)=\begin{cases}b & \text{if at least two agens are of type } t_b \text{ and } t_3\neq t_c\\ a & \text{otherwise}\end{cases}$$

It is easy to see that this modified SCF $\hat{f}$ still satisfies incentive compatibility and closure. I now show that the SCF $\hat{f}$ satisfies Bayesian monotonicity. Since every SCF trivially satisfies the convex-range property, this together with Proposition 3 in the previous section implies that $\hat{f}$ also satisfies mixed Bayesian monotonicity. Note that this example depicts an economy with common values and the SCF $\hat{f}$ specifies the best outcome in every state. So, to verify Bayesian monotonicity, I only need to show that for any deception β, whenever $f\circ\beta\not\approx f$, there exist agent $i\in N$, type $t_i\in\hat{T}_i$, and an SCF $y:\hat{T}_{-i}\to\Delta(A)$ with the following preference reversal:

$U_i(y\circ\beta_{-i}|t_i)>U_i(\hat{f}\circ\beta|t_i)$.

Before checking Bayesian monotonicity, I make the following observations: first, since $\hat{f}$ satisfies incentive compatibility, one can ignore unilateral deceptions from the profile of truth-telling. Second, there is no need to consider a deception of type t_c because, due to the construction of $\hat{f}$, type t_c obtain the best outcome a from $\hat{f}$ by announcing t_c regardless of the types of the other agents. So, I always assume $\beta_3(t_c)=t_c$ for any deception β. Given this consideration, the task here reduces to checking the following 10 cases of possible deceptions and finding the right preference reversal for each case.

Case 1: $\beta_i(t_a)=\beta_i(t_b)=t_b$ for each $i\in N$.

Choose agent 3 of type t_c and define $y(t_{-3})=a$ for all $t_{-3}\in\hat{T}_{-3}$. Since $U_3(f\circ\beta|t_c)=0$ and $U_3(y\circ\beta_{-3}|t_c)=1$, we have $U_3(y\circ\beta_{-3}|t_c)>U_3(\hat{f}\circ\beta|t_c)$.

Case 2: $\beta_i(t_a)=\beta_i(t_b)=t_a$ for each $i\in N$.

Choose agent 1 of type t_b as a test agent and set $y:\hat{T}_{-1}\to\Delta(A)$ as follows: for each $t_{-1}\in\hat{T}_{-1}$,

$$y(t_{-1})=\begin{cases}b & \text{if } t_2=t_3=t_a\\ a & \text{otherwise.}\end{cases}$$

Then, we compute the following:

$$U_1(\hat{f}\circ\beta|t_b)=(1-q)^2$$
$$U_1(y\circ\beta_{-1}|t_b)=q^2+2q(1-q)=1-(1-q)^2.$$

Since $q^2>1/2$, we obtain $U_1(y\circ\beta_{-1}|t_b)>U_1(\hat{f}\circ\beta|t_b)$.

Case 3-1: $\beta_i(t_a)=t_b$ and $\beta_i(t_b)=t_a$ for each $i\in N$.

Choose agent 1 of type t_b as a test agent and define $y:\hat{T}_{-1}\to\Delta(A)$ as follows: for each $t_{-1}\in\hat{T}_{-1}$,

$$y(t_{-1})=\begin{cases}a & \text{if } t_2=t_3=t_b\\ b & \text{otherwise.}\end{cases}$$

Then, we compute the following:

$$U_1(\hat{f}\circ\beta|t_b)=(1-q)^2$$
$$U_1(y\circ\beta_{-1}|t_b)=q^2+2q(1-q)=1-(1-q)^2.$$

Since $q^2>1/2$, we obtain $U_1(y\circ\beta_{-1}|t_b)>U_1(\hat{f}\circ\beta|t_b)$.

Case 3-2: $\beta_i(t_a)=t_b$ and $\beta_i(t_b)=t_a$ for each $i\in\{1,2\}$ and $\beta_3(t_b)=t_c$.

Choose agent 3 of type t_b as a test agent and define $y(t_{-3})=b$ for each $t_{-3}\in\hat{T}_{-3}$. We compute the following:

$$U_3(\hat{f}\circ\beta|t_b)=(1-q)^2,$$
$$U_3(y\circ\beta_{-3}|t_b)=q^2+2q(1-q)=1-(1-q)^2.$$

Since $q^2>1/2$, we obtain $U_3(y\circ\beta_{-3}|t_b)>U_3(\hat{f}\circ\beta|t_b)$.

Case 4: There exists $i\in N$ such that $\beta_i(t_i)=t_i$ for each $t_i\in\hat{T}_i$ and $\beta_j(t_a)=\beta_j(t_b)=t_b$ for each $j\in N\backslash\{i\}$.

Without loss of generality, we can set $i=3$. Then, choose agent 1 of type t_a as a test agent. Define $y:\hat{T}_{-1}\to\Delta(A)$ as follows: for any $t_{-i}\in\hat{T}_{-1}$,

$$y(t_{-1})=\begin{cases}b & \text{if } t_3=t_b\\ a & \text{otherwise.}\end{cases}$$

We compute the following:

$$U_1(\hat{f}\circ\beta|t_a)=q^2,$$
$$U_1(y\circ\beta_{-1}|t_a)=q^2+(1-q)^2+q(1-q).$$

This implies $U_1(y\circ\beta_{-1}|t_a)>U_1(\hat{f}\circ\beta|t_a)$.

Case 5: There exists $i\in N$ such that $\beta_i(t_i)=t_i$ for each $t_i\in\hat{T}_i$ and $\beta_j(t_a)=\beta_j(t_b)=t_a$ for each $j\in N\backslash\{i\}$.

Without loss of generality, we can set $i = 3$. Then, choose agent 1 of type t_b as a test agent. Define $y : \hat{T}_{-1} \to \Delta(A)$ as follows: for any $t_{-i} \in \hat{T}_{-1}$,

$$y(t_{-1}) = \begin{cases} b & \text{if } t_3 = t_b \\ a & \text{otherwise.} \end{cases}$$

We compute the following:

$$U_1(\hat{f} \circ \beta | t_b) = (1-q)^2,$$
$$U_1(y \circ \beta_{-1} | t_b) = q^2 + (1-q)^2 + q(1-q).$$

This implies $U_1(y \circ \beta_{-1}|t_b) > U_1(\hat{f} \circ \beta | t_b)$.

Case 6-1: There exists $i \in N$ such that $\beta_i(t_3) = t_3$ for each $t_3 \in \hat{T}_3$, $\beta_j(t_a) = t_b$ and $\beta_j(t_b) = t_a$ for each $j \in N \backslash \{i\}$.

Choose agent 3 of type t_b as a test agent and set $y(t_{-3}) = b$ for all $t_{-3} \in \hat{T}_{-3}$. Then we compute the following:

$$U_3(\hat{f} \circ \beta | t_b) = 2q(1-q),$$
$$U_3(y \circ \beta_{-3} | t_b) = q^2 + 2q(1-q).$$

This implies that $U_3(y \circ \beta_{-3}|t_b) > U_3(\hat{f} \circ \beta | t_b)$.

Case6-2: There exists $i \in \{1, 2\}$ such that $\beta_i(t_i) = t_i$ for each $t_i \in \hat{T}_i$, $\beta_j(t_a) = t_b$ and $\beta_j(t_b) = t_a$ for $j \in \{1, 2\} \backslash \{i\}$, and $\beta_3(t_a) = t_b$ and $\beta_3(t_b) = t_c$.

Choose agent 1 of type t_b as a test agent and define $y : \hat{T}_{-1} \to \Delta(A)$ as follows: for all $t_{-1} \in \hat{T}_{-1}$,

$$y(t_{-1}) = \begin{cases} b & \text{if } t_2 = t_b \text{ or } t_2 = t_a \text{ and } t_3 = t_c, \\ a & \text{otherwise.} \end{cases}$$

We compute the following:

$$U_1(\hat{f} \circ \beta | t_b) = 2q(1-q),$$
$$U_1(y \circ \beta_{-1} | t_b) = 1.$$

This implies that $U_1(y \circ \beta_{-i}|t_b) > U_1(\hat{f} \circ \beta | t_b)$.

Case6-3: There exists $i \in \{1, 2\}$ such that $\beta_i(t_i) = t_i$ for each $t_i \in \hat{T}_i$, $\beta_j(t_a) = t_b$ and $\beta_j(t_b) = t_a$ for $j \in \{1, 2\} \backslash \{i\}$, and $\beta_3(t_a) = t_c$ and $\beta_3(t_b) = t_a$.

Choose agent 1 of type t_b as a test agent and define $y : \hat{T}_{-1} \to \Delta(A)$ as follows: for all $t_{-1} \in \hat{T}_{-1}$,

$$y(t_{-1}) = \begin{cases} b & \text{if } t_2 = t_b \text{ or } t_2 = t_a \text{ and } t_3 = t_a \\ a & \text{otherwise.} \end{cases}$$

We then compute the following:

$$U_1(\hat{f} \circ \beta | t_b) = q(1-q),$$
$$U_1(y \circ \beta_{-1} | t_b) = q^2 + 2q(1-q).$$

This implies that $U_1(y \circ \beta_{-1}|t_b) > U_1(\hat{f} \circ \beta | t_b)$.

Case6-4: There exists $i \in \{1, 2\}$ such that $\beta_i(t_i) = t_i$ for each $t_i \in \hat{T}_i$, $\beta_j(t_a) = t_b$ and $\beta_j(t_b) = t_a$ for $j \in \{1, 2\} \backslash \{i\}$, and $\beta_3(t_a) = \beta_3(t_b) = t_c$.

Choose agent 1 of type t_b as a test agent and define $y : \hat{T}_{-1} \to \Delta(A)$ as follows: for each $t_{-1} \in \hat{T}_{-1}$,

$$y(t_{-1}) = \begin{cases} b & \text{if } t_2 = t_b \text{ or } t_2 = t_a \text{ and } t_3 = t_c, \\ a & \text{otherwise.} \end{cases}$$

We then compute the following:

$$U_1(\hat{f} \circ \beta | t_b) = (1-q)^2,$$
$$U_1(y \circ \beta_{-1} | t_b) = q^2 + q(1-q) + (1-q)^2.$$

This implies that $U_1(y \circ \beta_{-1}|t_b) > U_1(\hat{f} \circ \beta | t_b)$.

For every possible deception, I found a test agent who has the right preference reversal. Hence, $\hat{f}$ satisfies Bayesian monotonicity.

7. Sufficiency for mixed Bayesian implementation

In this section, I discuss the sufficient conditions for mixed Bayesian implementation. First, I prepare some definitions. For each SCF $f \in \mathbb{F}$, define

$$Y_i[f] \equiv \left\{ y_i : T_{-i} \to \Delta^*(A) | U_i(f|\tilde{t}_i) \geq U_i(y_i|\tilde{t}_i) \ \forall \tilde{t}_i \in T_i^* \right\},$$

where $\Delta^*(A)$ is a finite subset of $\Delta(A)$. $Y_i[f]$ describes the set of SCFs that are at least as bad as the SCF f for agent i, regardless of his type. Since T_{-i} and $\Delta^*(A)$ are finite, $Y_i[f]$ becomes a finite set.[18]

I introduce a condition on SCSs, which plays an important role in the sufficiency results (Theorems 1–3) later.

Definition 9. An SCS F satisfies the **no-worst-rule** condition (NWR) if, for each $f \in F$, $i \in N$, $t_i \in T_i^*$, and $\psi_i \in \Delta(T_{-i} \times T_{-i})$, there exist two SCFs $y_i[f; t_i, \psi_i], y_i'[f; t_i, \psi_i] \in Y_i[f]$ such that

$$\sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) u_i(y'_i[f; t_i, \psi_i](t'_{-i}); (t_i, t_{-i}))$$
$$> \sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) u_i(y_i[f; t_i, \psi_i](t'_{-i}); (t_i, t_{-i})).$$

Remark. By NWR, I require that for any belief that type t_i has about both t_{-i} and how types t_{-i} pretend to be t'_{-i} in the direct revelation mechanism, there be no total indifference among all SCFs within $Y_i[f]$. This NWR is an incomplete information analogue of the conditional-no-total-indifference condition of Bergemann and Morris (2011) and moreover it is extended so as to take care of social choice "sets".

To illustrate the permissiveness of NWR, I will revisit the example discussed in the previous section.

Example 2 (*Permissiveness of NWR*). To illustrate the permissiveness of NWR, we slightly modify the previous example in Section 6 by adding the third alternative c to $\{a, b\}$ so that $A = \{a, b, c\}$ and for each $i \in N$ and $t \in T$, $u_i(c; t) = 0$. We shall show that the majoritarian SCF f, which was introduced in the previous section, indeed satisfies NWR. As we will see below, what is essential for the argument is not that there is a common alternative c but that each agent has a different alternative like c, which depends also on his belief.

Since all agents are symmetric, we focus on agent 1. Consider type t_a of agent 1 and fix $\psi_1 \in \Delta(T_{-1} \times T_{-1})$. Define

$$y_1[f; t_a, \psi_1] = \begin{cases} a & \text{with probability } p_a, \\ b & \text{with probability } p_b, \\ c & \text{with probability } p_c, \end{cases}$$

$$y'_1[f; t_a, \psi_1] = \begin{cases} a & \text{with probability } p'_a, \\ b & \text{with probability } p'_b, \\ c & \text{with probability } p'_c, \end{cases}$$

to be the constant SCFs where $p = (p_a, p_b, p_c), p' = (p'_a, p'_b, p'_c) \in \Delta(A)$. We will choose particular values of p and p' later. Since $y_1[f; t_a, \psi_1]$ and $y'_1[f; t_a, \psi_1]$ are constant SCFs and the majoritarian SCF f chooses the best outcome in every state, we know that $y_1[f; t_a, \psi_1], y'_1[f; t_a, \psi_1] \in Y_1[f]$. Define

$$q_a = \sum_{t'_{-1} \in T_{-1}} \left\{ \psi_1(t'_{-1}, t_a, t_a) + \psi_1(t'_{-1}, t_a, t_b) + \psi_1(t'_{-1}, t_b, t_a) \right\},$$

[18] In the Appendix, I will extend the range of y_i from $\Delta^*(A)$ to $\Delta(A)$. Thus, the finiteness of $Y_i[f]$ is only made for simplifying the argument.

$$q_b = \sum_{t'_{-1} \in T_{-1}} \psi_1(t'_{-1}, t_b, t_b),$$

where $q_a + q_b = 1$. Then, we compute the following:

$$\sum_{t'_{-1}, t_{-1}} \psi_1(t'_{-1}, t_{-1}) u_1(y_1[f; t_a; \psi_1](t'_{-1}); t_a, t_{-1}) = p_a q_a + p_b q_b,$$

$$\sum_{t'_{-1}, t_{-1}} \psi_1(t'_{-1}, t_{-1}) u_1(y'_1[f; t_a; \psi_1](t'_{-1}); t_a, t_{-1}) = p'_a q_a + p'_b q_b.$$

By choosing $p'_a > p_a$ and $p'_b > p_b$ ($p'_c < p_c$), we obtain the right preference reversal regardless of the values of q_a and q_b. We can make essentially the same argument for type t_b. Hence, NWR can be easily satisfied in this modified example.

Since T_i is finite, I define $\{t_i^\ell\}_{\ell=1}^\infty$ as a countable support of $\Delta(T_{-i})$. Similarly, since T_{-i} is finite, one can find $\{\psi_i^k\}_{k=1}^\infty$ as a countable support of $\Delta(T_{-i} \times T_{-i})$.

Since F satisfies NWR, for each $f \in F$ and $i \in N$, I define the uniform SCF $\bar{y}_i[f]$ as follows: there exist $\delta, \eta \in (0, 1)$ such that

$$\bar{y}_i[f] \equiv \frac{(1-\delta)(1-\eta)}{2} \sum_{\ell=1}^\infty \eta^{\ell-1} \times \sum_{k=1}^\infty \delta^{k-1} \left\{ y'_i[f; t_i^\ell, \psi_i^k] + y_i[f; t_i^\ell, \psi_i^k] \right\}.$$

I use this uniform SCF $\bar{y}_i[f]$ in the canonical mechanism I propose later. I also use the following result later.

Lemma 1. *Suppose that an SCS F satisfies NWR. Then, for all $i \in N$, $t_i \in T_i^*$, and $\phi_i \in \Delta(T_{-i})$, there exist two lotteries (or constant SCFs) $\alpha_i[t_i, \phi_i], \alpha'_i[t_i, \phi_i] \in \Delta(A)$ such that*

$$\sum_{t_{-i}} \phi_i(t_{-i}) u_i(\alpha'_i[t_i, \phi_i]; (t_i, t_{-i})) > \sum_{t_{-i}} \phi_i(t_{-i}) u_i(\alpha_i[t_i, \phi_i]; (t_i, t_{-i})).$$

Remark. The property stated in this lemma is slightly stronger than the version of the no-total-indifference condition used by SV (2005, 2010).

Proof. The proof is in the Appendix. ■

Following the previous argument, I denote by $\{t_i^\ell\}_{\ell=1}^\infty$ its countable support of $\Delta(T_i)$ and by $\{\phi_i^k\}_{k=1}^\infty$ its countable support of $\Delta(T_{-i})$, respectively. For each $i \in N$, I define the uniform lottery $\bar{\alpha}_i \in \Delta(A)$ as follows: there exist $\delta, \eta \in (0, 1)$ such that

$$\bar{\alpha}_i \equiv \frac{(1-\delta)(1-\eta)}{2} \sum_{\ell=1}^\infty \eta^{\ell-1} \sum_{k=1}^\infty \delta^{k-1} \left\{ \alpha'_i[t_i^\ell, \phi_i^k] + \alpha_i[t_i^\ell, \phi_i^k] \right\}.$$

Finally, I define

$$\bar{\alpha} \equiv \frac{1}{n} \sum_{i \in N} \bar{\alpha}_i.$$

I use this uniform lottery $\bar{\alpha}$ in the canonical mechanism I propose below and I am ready to provide a sufficiency result for mixed Bayesian implementation.

Theorem 1. *Suppose that there are at least three agents ($n \geq 3$). If an SCS F satisfies incentive compatibility, mixed Bayesian monotonicity, closure, and NWR, then it is mixed Bayesian implementable.*

Remark. The proof here builds upon the canonical mechanism proposed in Proposition 1 of Oury and Tercieux (2012). Theorem 1 of SV (2010) uses a mechanism whose message space consists of five components and each component is almost the same as that in the canonical mechanism proposed below. In this sense, the mechanism below is no more complex than that of SV (2010).

Proof. We prove this by constructing an implementing mechanism $\Gamma = (M, g)$. Each agent i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4, m_i^5) \in M_i$ where $m_i^1 \in T_i^*$; $m_i^2 \in F$; $m_i^3 \in \mathbb{N} = \{1, 2, \ldots\}$; $m_i^4 = \{m_i^4[f]\}_{f \in F}$ where $m_i^4[f] \in Y_i[f]$ for each $f \in F$; and $m_i^5 \in \Delta^*(A)$, which is a finite subset of $\Delta(A)$. Since the type space T, the SCS F, each $Y_i[f]$, and $\Delta^*(A)$ are all finite, each message space M_i is made countable.

The outcome $g(m)$ is determined by the following rules:

Rule 1: If there exists $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for all $j \in N$, then $g(m) = f(m^1)$.

Rule 2: If there exist $i \in N$ and $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for all $j \neq i$ and $m_i^2 \not\approx f$ or $m_i^3 > 1$, then

$$g(m) = \begin{cases} m_i^4[f](m_{-i}^1) & \text{with probabiltiy } m_i^3/(m_i^3+1) \\ \bar{y}_i[f](m_{-i}^1) & \text{with probability } 1/(m_i^3+1) \end{cases}$$

where $\bar{y}_i[f]$ is the uniform SCF defined previously.

Rule 3: In all other cases,

$$g(m) = \begin{cases} m_1^5 & \text{with probability } m_1^3/n(m_1^3+1) \\ m_2^5 & \text{with probability } m_2^3/n(m_2^3+1) \\ \vdots & \vdots \\ m_n^5 & \text{with probability } m_n^3/n(m_n^3+1) \\ \bar{\alpha} & \text{with the remaining probability} \end{cases}$$

where $\bar{\alpha}$ is the uniform lottery over A, as defined previously.

The proof is completed by a series of claims below.

Claim 1. *For any Bayesian Nash equilibrium $\sigma \in BNE^{\Gamma(T)}$, $m \in M$, and $t \in T^*$: $m \in supp(\sigma(\cdot|t)) \Rightarrow m_j^3 = 1$ for each $j \in N$.*

Proof of Claim 1. Fix a Bayesian Nash equilibrium $\sigma \in BNE^{\Gamma(T)}$. We focus on type t_i of agent i. Suppose by way of contradiction that $\text{supp}(\sigma_i(\cdot|t_i))$ contains a message m_i such that $m_i^3 > 1$. We fix such m_i and partition the messages of all agents but i as follows: for each $t'_{-i} \in T_{-i}$,

$$M_{-i}^*(t'_{-i}) = \left\{ m_{-i} \middle| \; \exists f \in F \text{ s.t. } m_j^2 \approx f \text{ and } m_j^3 = 1 \text{ for all } j \neq i \text{ and } m_{-i}^1 = t'_{-i} \right\},$$

and

$$\hat{M}_{-i} = \left\{ m_{-i} \middle| \; m_j^3 > 1 \text{ for some } j \in N \backslash \{i\} \text{ or } \exists j, k \in N \backslash \{i\} \text{ s.t. } m_j^2 \not\approx m_k^2 \right\}.$$

If

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i}) > 0,$$

then, we can set

$$\phi_i(t_{-i}) = \gamma \cdot \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i})$$

for any $t_{-i} \in T_{-i}$ where the constant γ is chosen so that $\sum_{t_{-i}} \phi_i(t_{-i}) = 1$. Since F satisfies NWR, by Lemma 1, we know that there exists $\hat{m}_i^5 \in \Delta(A)$ such that

$$\sum_{t_{-i}} \phi_i(t_{-i}) u_i(\hat{m}_i^5; (t_i, t_{-i})) > \sum_{t_{-i}} \phi_i(t_{-i}) u_i(\bar{\alpha}; (t_i, t_{-i})).$$

Define $\hat{m}_i$ to be the same as m_i except that m_i^5 is replaced by $\hat{m}_i^5$ defined above and $\hat{m}_i^3 = m_i^3 + 1$. Then, the above inequality allows us to conclude the following:

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i})$$
$$\times \left[u_i(g(\hat{m}_i, m_{-i}); (t_i, t_{-i})) - u_i(g(m_i, m_{-i}); (t_i, t_{-i}))\right] > 0.$$

This implies that $U_i(g \circ (\hat{m}_i, \sigma_{-i})|t_i) > U_i(g \circ (m_i, \sigma_{-i})|t_i)$. Thus, $\hat{m}_i$ would be an even better response to σ_{-i} than m_i. This contradicts to our hypothesis that $\sigma \in BNE^{\Gamma(T)}$.

For each $t'_{-i} \in T_{-i}$, if

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i} \in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|t_{-i}) > 0,$$

then, we can set

$$\psi_i(t'_{-i}, t_{-i}) \equiv \gamma \cdot \pi_i[t_i](t_{-i}) \sum_{m_{-i} \in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|t_{-i}),$$

for any $t_{-i} \in T_{-i}$ where the constant γ is chosen so that $\sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) = 1$. Since F satisfies NWR, there exists $\hat{m}^4_i[f] \in Y_i[f]$ such that

$$\sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) u_i(\hat{m}^4_i[f](t'_{-i}); (t_i, t_{-i}))$$
$$> \sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) u_i(\bar{y}_i[f](t'_{-i}); (t_i, t_{-i})).$$

Define $\hat{m}_i$ to be the same as m_i except that $m^4_i[f]$ is replaced by $\hat{m}^4_i[f]$ defined above and $\hat{m}^3_i = m^3_i + 1$. Then, the above inequality allows us to conclude the following:

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i})$$
$$\times \left[u_i(g(\hat{m}_i, m_{-i}); (t_i, t_{-i})) - u_i(g(m_i, m_{-i}); (t_i, t_{-i}))\right] > 0.$$

This implies that $U_i(g \circ (\hat{m}_i, \sigma_{-i})|t_i) > U_i(g \circ (m_i, \sigma_{-i})|t_i)$. Thus, $\hat{m}_i$ would be an even better response to σ_{-i} than m_i. This contradicts our hypothesis that $\sigma \in BNE^{\Gamma(T)}$. ■

Claim 2. *For any equilibrium $\sigma \in BNE^{\Gamma(T)}$, there exists $f \in F$ such that $m^2_j \approx f$ for each $j \in N$, $t_j \in T^*_j$, and $m_j \in supp(\sigma_j(\cdot|t_j))$.*

Proof of Claim 2. Fix an equilibrium $\sigma \in BNE^{\Gamma(T)}$. By the previous claim, we have that for any $t \in T^*$ and $m \in \text{supp}(\sigma(\cdot|t))$, $m^3_j = 1$ for each $j \in N$. Suppose by way of contradiction that there exist $t \in T^*$ and $i \in N$ such that $\text{supp}(\sigma_i(\cdot|t_i))$ contains a message m_i such that $m^3_i = 1$ and $m^2_i \not\approx m^2_k$ for some $k \neq i$. Then, we fix such m_i and only need to consider the following two cases:

Case 1: there exist $j, k \in N\backslash\{i\}$ with $j \neq k$ such that $m^2_i \not\approx m^2_j$, $m^2_j \not\approx m^2_k$, and $m^2_k \not\approx m^2_i$; or

Case 2: there exists $f \in F$ such that $m^2_j \approx f$ for each $j \neq i$

By our hypothesis, either Case 1 or Case 2 occurs with probability one. We focus on agent i of type t_i throughout. First, we assume Case 1 applies. We consider the following message profiles of all agents but i:

$$\hat{M}_{-i} = \left\{m_{-i} \middle| \exists j, k \in N\backslash\{i\} \text{s.t.} m^2_i \not\approx m^2_j,\ m^2_j \not\approx m^2_k, \text{ and } m^2_k \not\approx m^2_i\right\}.$$

If

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i}) > 0,$$

then, we can set

$$\phi_i(t_{-i}) \equiv \gamma \cdot \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i}),$$

for any $t_{-i} \in T_{-i}$ where the constant γ is chosen so that $\sum_{t_{-i}} \phi_i(t_{-i}) = 1$. Since F satisfies NWR, by Lemma 1 and continuity of expected utility, there exists $\hat{m}^5_i \in \Delta^*(A)$ such that

$$\sum_{t_{-i}} \phi_i(t_{-i}) u_i(\hat{m}^5_i; (t_i, t_{-i})) > \sum_{t_{-i}} \phi_i(t_{-i}) u_i(\bar{\alpha}; (t_i, t_{-i})).$$

Next we assume that Case 2 applies. We consider the following messages of all agents but i: for each $t'_{-i} \in T_{-i}$,

$$M^*_{-i}(t'_{-i}) = \left\{m_{-i} \middle| \exists f \in F \text{ s.t. } m^2_j \approx f \forall j \neq i \text{ and } m^1_{-i} = t'_{-i}\right\}.$$

For any $t'_{-i} \in T_{-i}$, if

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i} \in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|t_{-i}) > 0,$$

then, we can set

$$\psi_i(t'_{-i}, t_{-i}) \equiv \gamma \cdot \pi_i(t_i)[t_{-i}] \sum_{m_{-i} \in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|t_{-i}),$$

for any $t_{-i} \in T_{-i}$ where the constant γ is chosen so that $\sum_{t_{-i}, t'_{-i}} \psi_i(t'_i, t_i) = 1$. Since F satisfies NWR, there exists $\hat{m}^4_i[f] \in Y_i[f]$ such that

$$\sum_{t_{-i}, t'_{-i}} \psi_i(t'_{-i}, t_{-i}) u_i(\hat{m}^4_i[f](t'_{-i}); (t_i, t_{-i}))$$
$$> \sum_{t_{-i}, t'_{-i}} u_i(\bar{y}_i[f](t'_{-i}); (t_i, t_{-i})),$$

Define $\hat{m}_i$ to be the same as m_i except that $m^4_i[f]$ is replaced by $\hat{m}^4_i[f]$; m^5_i is replaced by $\hat{m}^5_i$; and $\hat{m}^3_i = m^3_i + 1$. Then, we conclude the following:

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i})$$
$$\times \left[u_i(g(\hat{m}_i, m_{-i}); (t_i, t_{-i})) - u_i(g(m_i, m_{-i}); (t_i, t_{-i}))\right] > 0.$$

This implies that $U_i(g \circ (\hat{m}_i, \sigma_{-i})|t_i) > U_i(g \circ (m_i, \sigma_{-i})|t_i)$. Thus, $\hat{m}_i$ would be an even better response to σ_{-i} than m_i. This contradicts our hypothesis that $\sigma \in BNE^{\Gamma(T)}$. ■

Claim 3. *For each $f \in F$, there exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for all $j \in N$ and $t_j \in T_i$:*

$$m_j \in supp(\sigma_j(\cdot|t_j)) \Rightarrow m_j = (t_j, f, 1, m^4_j, m^5_j)$$

for some $m^4_j \in M^4_j$ and $m^5_j \in M^5_j$.

Proof of Claim 3. Fix $f \in F$. For each $j \in N$, define $\sigma_j : T_j \to M_j$ satisfying the following properties: for each $t_j \in T^*_j$ and $m_j \in M_j$, whenever $m_j \in \text{supp}(\sigma^1_j(\cdot|t_j))$,

1. $m^1_j = t_j$;
2. $m^2_j = f$; and
3. $m^3_j = 1$.

Fix $i \in N$ and $t_i \in T_i$. It only remains to show that $\sigma_i(t_i)$ is a best response of type t_i to σ_{-i}. Observe that m^5_i is irrelevant for the resulting outcome given σ_{-i}. If type t_i deviates from $\sigma_i(t_i)$ to some $\hat{m}_i$ where $\hat{m}^2_i \not\approx f$ or $\hat{m}^3_i > 1$, he can induce Rule 2. Then, there is always positive probability that $\bar{y}_i[f]$ is chosen and since F satisfies NWR and all agents are truthful by our hypothesis, type t_i can be worse off by this deviation than playing $\sigma_i(t_i)$. Thus, the only profitable deviation type t_i can possibly have is to change m_i^1 but change neither m_i^2 nor m_i^3. However, since all agents are truthful by our hypothesis and f satisfies incentive compatibility, this cannot be profitable. Thus, $\sigma_i(t_i)$ is a best response for type t_i to σ_{-i}. ■

Claim 4. *For each SCF $f \in F$ and deception β, if there exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for each $t \in T^*$, $m \in supp\ (\sigma(\cdot|t))$, and $j \in N$,*

$$m_j = (t_j', f, 1, m_j^4, m_j^5)$$

for some $(m_j^4, m_j^5) \in M_j^4 \times M_j^5$ and $t_j' \in \beta_j(t_j)$,

then, $f \circ \beta \in F$.

Proof of Claim 4. Fix an SCF $f \in F$, a deception β, and an equilibrium $\sigma \in BNE^{\Gamma(T)}$ as defined in the statement of the claim. We argue by contradiction. Suppose $f \circ \beta \notin F$. Then, by mixed Bayesian monotonicity, we have that for every collection of $\{\psi_k\}_{k \in N}$ with each function $\psi_k : T_k \to \Delta(T_k)$ such that for each $k \in N$ and $t_k, t_k' \in T_k$,

$$\psi_k(t_k'|t_k) > 0 \Leftrightarrow t_k' \in \beta_k(t_k),$$

there exist $i \in N$, $t_i \in T_i^*$, and a function $y^* : T_{-i} \to \Delta(A)$ such that

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{t_{-i}'} \psi_{-i}(t_{-i}', t_{-i}) \times \left[u_i(y^*(t_{-i}'), (t_i, t_{-i})) - \sum_{t_i'} \psi_i(t_i'|t_i) u_i(f(t_i', t_{-i}'); (t_i, t_{-i})) \right] > 0,$$

where $\psi_{-i}(t_{-i}'|t_{-i}) = \prod_{j \neq i} \psi_j(t_j'|t_j)$ and for all $\tilde{t}_i \in T_i^*$,

$$\sum_{t_{-i}} \pi_i(\tilde{t}_i)[t_{-i}] u_i(f(\tilde{t}_i, t_{-i}); (\tilde{t}_i, t_{-i})) \geq \sum_{t_{-i}} \pi_i(\tilde{t}_i)[t_{-i}] u_i(y^*(t_{-i}), (\tilde{t}_i, t_{-i})).$$

For each $j \in N$ and $t_j' \in T_j$, define

$$M_j(t_j') \equiv \left\{ m_j \in M_j \middle| \ m_j^1 = t_j' \right\},$$

and for each $j \in N$, $t_j \in T_j$, and $t_j' \in T_j$, define

$$\psi_j(t_j'|t_j) \equiv \sum_{m_j \in M_j(t_j')} \sigma_j(m_j|t_j).$$

In what follows, we focus on agent i of type t_i as identified in the condition of mixed Bayesian monotonicity. Assume by our hypothesis that $\text{supp}(\sigma_i(\cdot|t_i))$ contains a message m_i as such. For each $t_{-i}' \in T_{-i}$, define

$$M_{-i}(t_{-i}') = \prod_{j \neq i} M_j(t_j').$$

For each $t_{-i}, t_{-i}' \in T_{-i}$, define

$$\psi_{-i}(t_{-i}'|t_{-i}) \equiv \sum_{m_{-i} \in M_{-i}(t_{-i}')} \sigma_{-i}(m_{-i}|t_{-i}).$$

Then, by mixed Bayesian monotonicity, we can set $\hat{m}_i^4[f] = y^* \in Y_i[f]$ so that

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{t_{-i}'} \psi_{-i}(t_{-i}', t_{-i}) \times \left[u_i(y^*(t_{-i}'), (t_i, t_{-i})) - \sum_{t_i'} \psi_i(t_i'|t_i) u_i(f(t_i', t_{-i}'); (t_i, t_{-i})) \right] > 0,$$

Define $\hat{m}_i$ to be the same as m_i except that $m_i^4[f]$ is replaced by $\hat{m}_i^4[f] = y^*$ and $\hat{m}_i^3$ is chosen sufficiently large. Then, the above inequality allows us to conclude the following:

$$\sum_{t_{-i}} \pi_i(t_i)[t_{-i}] \sum_{m_{-i}} \sigma_{-i}(m_{-i}|t_{-i}) \times \left[u_i(g(\hat{m}_i, m_{-i}); (t_i, t_{-i})) - u_i(g(m_i, m_{-i}); (t_i, t_{-i})) \right] > 0.$$

That is, $U_i(g \circ (\hat{m}_i, \sigma_{-i})|t_i) > U_i(g \circ (m_i, \sigma_{-i})|t_i)$. This implies that, given σ_{-i} of the other agents' equilibrium strategies, $\hat{m}_i$ can increase his payoff by announcing $\hat{m}_i^3$ sufficiently large, choosing $\hat{m}_i^4[f] = y^*$, and, as a result, inducing Rule 2. That is, $\hat{m}_i$ is an even better response to σ_{-i} than m_i. This contradicts our hypothesis that $\sigma \in BNE^{\Gamma(T)}$. ■

All four claims we have established above imply that (1) for any $f \in F$, there exists $\sigma \in BNE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for any $\sigma \in BNE^{\Gamma(T)}$, we have that $g \circ \sigma \in F$. Thus, we complete the proof of the theorem. ■

Next, I extend the previous result to the case of two agents. The case of two agents is prevalent in the literature of contract theory and industrial organization because one can interpret a bilateral contracting problem as a mechanism design problem with two agents. For the case of two agents, Dutta and Sen (1994) establish the following necessary condition: for all $f^1, f^2 \in F$ with $f^1 \not\approx f^2$, we must have $Y_1[f^2] \cap Y_2[f^1] \neq \emptyset$. To obtain a sufficiency result, I slightly strengthen this necessary condition in a similar fashion as NWR is defined:

Definition 10. An SCS F satisfies the **intersection property** if, for each $i \in \{1, 2\} = N$, $f^i, f^j \in F$ with $f^i \not\approx f^j$, $t_i \in T_i^*$, and $\psi_i \in \Delta(T_{-i} \times T_{-i})$, there exist $y_i[f^j; t_i, \psi_i] \in Y_i[f^j]$ and $y_i[f^i, f^j; t_i, \psi_i] \in Y_i[f^j] \cap Y_j[f^i]$ such that

$$\sum_{t_{-i}, t_{-i}'} \psi_i(t_{-i}', t_{-i}) \left[u_i(y_i[f^j; t_i, \psi_i](t_{-i}'); (t_i, t_{-i})) - u_i(y_i[f^i, f^j; t_i, \psi_i](t_{-i}'); (t_i, t_{-i})) \right] > 0.$$

I establish the following theorem for the case of two agents:

Theorem 2. *Let $N = \{1, 2\}$ be the set of agents. If an SCS F satisfies incentive compatibility, mixed Bayesian monotonicity, closure, NWR and the intersection property, it is mixed Bayesian implementable.*

Proof. The proof is in the Appendix. ■

Finally, I adapt Theorems 1 and 2 to the case of social choice "functions". Most importantly, I propose a single mechanism that works simultaneously for the case of two agents as well as more than two agents.

Theorem 3. *If an SCF f satisfies incentive compatibility, Bayesian monotonicity, and NWR, it is mixed Bayesian implementable.*

Proof. The proof is in the Appendix. ■

8. Mixed Nash implementation

In this section, I investigate the conditions for *mixed Nash implementation*, which restrict its analysis to a *complete information* environment. The objective of this section is to appropriately adapt our Theorem 1 of mixed Bayesian implementation to complete information environments. I start from the formal definition of complete information.

Definition 11. $(T_i, \pi_i)_{i\in N}$ is said to be a **complete information** type space if every $t \in T$ is a belief-closed subspace.

In the rest of the section, I assume that $(T_i^*, \pi_i^*)_{i\in N}$ is a belief-closed subspace in a complete information type space $(T_i, \pi_i)_{i\in N}$. The solution concept I adopt here is *Nash equilibrium*. A strategy profile $\sigma \in \Sigma$ is a Nash equilibrium of the game $\Gamma(T)$ if, for every $i \in N,\ t \in T$, and $\sigma_i' \in \Sigma_i$,

$$u_i(g(\sigma(t)); t) \geq u_i(g(\sigma_i'(t_i), \sigma_{-i}(t_{-i})); t),$$

where $t = (t_i, t_{-i})$. Let $NE^{\Gamma(T)}$ denote the set of Nash equilibria of the game $\Gamma(T)$. I now define *Nash implementation*.

Definition 12. An SCS F is **mixed Nash implementable** if there exists a mechanism Γ such that $g \circ NE^{\Gamma(T)} \approx F$. More specifically, this requirement can be decomposed into the following two properties: (1) for any SCF $f \in F$, there exists $\sigma \in NE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for any $\sigma \in NE^{\Gamma(T)}$, there exists $\hat{f} \in F$ such that $g \circ \sigma \approx \hat{f}$.

Remark. This definition is weaker than the one used by Maskin (1999) in his mixed-strategy Nash implementation. An SCS F is mixed Nash implementable "in the sense of Maskin" if the following two conditions hold: (I) for every $f \in F$, there exists $\sigma \in NE^{\Gamma(T)}$ such that for any $t \in T^*$ and $m \in \text{supp}(\sigma(\cdot|t))$, $g(m) = f(t)$; and (II) for any $\sigma \in NE^{\Gamma(T)}$, $t \in T^*$, and $m \in \text{supp}(\sigma(\cdot|t))$, $g(m) \in F(t)$. See also SV (2010, Section 5.1, pp. 783–4) for further discussion.

Maskin (1999) proposes a monotonicity condition for (mixed) Nash implementation.

Definition 13. An SCS F satisfies **Maskin monotonicity** if, for any $f \in F$ and $t, t' \in T^*$, whenever $f(t') \notin F(t)$, there exist $i \in N$ and $\alpha \in \Delta(A)$ such that

$$u_i(\alpha; t) > u_i(f(t'); t) \text{and} u_i(f(t'); t') \geq u_i(\alpha; t').$$

In what follows, I establish three preliminary results (Lemma 2 and Propositions 5 and 6) which allow us to unify Bayesian implementation and Nash implementation. I start from the result below, which shows an equivalence between Bayesian monotonicity and Maskin monotonicity in complete information environments where incentive compatibility is satisfied.

Proposition 5. *Suppose that $(T_i, \pi_i)_{i\in N}$ is a complete information type space and an SCS F satisfies incentive compatibility. Then, F satisfies Maskin monotonicity if and only if it satisfies Bayesian monotonicity.*

Proof. The proof is in the Appendix. ■

I show that "every" SCS satisfies the convex range property.

Lemma 2. *Suppose $(T_i, \pi_i)_{i\in N}$ is a complete information type space. Then, **every** SCS F satisfies the convex range property.*

Proof. Fix $f \in F$ and a collection of "single-valued" deceptions $\{\beta^\lambda\}_{\lambda\in\Lambda}$ such that $f \circ \beta^\lambda \in F$ for each $\lambda \in \Lambda$. Assume that there exists (not necessarily single-valued) deception $\tilde{\beta}$ such that $\tilde{\beta}(t) = \bigcup_\lambda \beta^\lambda(t)$ for every $t \in T^*$. Since every $t \in T^*$ is a belief-closed subspace under complete information, we can conclude that $f \circ \tilde{\beta}(t) \in F(t)$ for every $t \in T^*$, which implies $f \circ \tilde{\beta} \in F$. ■

The next result shows an equivalence between Maskin monotonicity and mixed Bayesian monotonicity.

Corollary 1. *Suppose that $(T_i, \pi_i)_{i\in N}$ is a complete information type space and an SCS F satisfies incentive compatibility. Then, F satisfies Maskin monotonicity if and only if it satisfies mixed Bayesian monotonicity.*

Proof. This follows from Proposition 5 and Lemma 2 ■

The next result shows that incentive compatibility becomes a vacuous constraint in complete information environments where there are at least three agents.

Proposition 6. *Suppose that $(T_i, \pi_i)_{i\in N}$ is a complete information type space and there are at least three agents, i.e., $n \geq 3$. Then, for every SCS F, there exists an SCS $\hat{F} \approx F$ such that $\hat{F}$ satisfies interim incentive compatibility.*

Proof. When there are at least three agents, any unilateral deviation from the truth telling can be detected. Thus, incentive compatibility trivially holds. ■

I introduce a significantly weaker version of NWR:

Definition 14. An SCS F satisfies the **no-worst-alternative** condition (henceforth, NWA) if, for each $f \in F$ and $i \in N$, there exists an SCF $y_i[f] : T \to \Delta(A)$ such that

$$u_i(f(t); t) > u_i(y_i[f](t); t)$$

for all $t \in T^*$.

Remark. This is what Cabrales and Serrano (2011) proposed as "the no-worst-alternative" condition. In words, NWA says that the SCS never assign the worst outcome to any agent at any state.

I obtain the following sufficiency result for mixed Nash implementation:

Theorem 4. *Suppose that $(T_i, \pi_i)_{i\in N}$ is a complete information type space and there are at least three agents, i.e., $n \geq 3$. If an SCS F satisfies Maskin monotonicity and NWA, there exists an SCS $\hat{F} \approx F$ such that $\hat{F}$ is mixed Nash implementable.*

Proof. The proof is in the Appendix. ■

9. Conclusion

The current paper identifies the no-worst rule condition (NWR) as a mild condition under which mixed Bayesian implementation is fully characterized. In so doing, I assess, by means of examples, the implications of the common prior assumption in mixed Bayesian implementation. I also cover the case of two agents, social choice functions, and Nash implementation under complete information. The main contribution of this paper is to propose a unification of the literature of Bayesian implementation and Nash implementation in general environments including non-economic ones.

I conclude this paper with what remains to be done. To appreciate better this paper's results, I believe providing a single example that touches upon all the key ingredients simultaneously is desirable. But I consider this endeavor challenging at this stage so that I want to leave it as a future work.

Appendix

In this section, I provide all the omitted proofs of the results of the paper.

A.1. Proof of Proposition 3

By definition, it is clear that mixed Bayesian monotonicity implies Bayesian monotonicity.

Now, we shall show the converse. Suppose that F satisfies Bayesian monotonicity. Fix $f \in F$. Consider a "non" single-valued deception $\tilde{\beta}$ such that $f \circ \tilde{\beta} \notin F$ and a collection of $\{\psi_k\}_{k \in N}$ with each function $\psi_k : T_k \to \Delta(T_k)$ such that for each $k \in N$ and $t_k \in T_k$,

$$\psi_k(t_k'|t_k) > 0 \Leftrightarrow t_k' \in \tilde{\beta}_k(t_k).$$

For this particular $\tilde{\beta}$, one can find a collection of single-valued deceptions $\{\beta^\lambda\}_{\lambda \in \Lambda}$ together with Λ being the index set such that (1) $\tilde{\beta} = \bigcup_{\lambda \in \Lambda} \beta^\lambda$ and (2) there do not exist $\lambda, \lambda' \in \Lambda$ with $\lambda \neq \lambda'$ for which $\beta^\lambda(\tilde{t}) = \beta^{\lambda'}(\tilde{t})$ for some $\tilde{t} \in T^*$. Since F satisfies the convex range property, we can find a "single-valued" deception β^0 such that $\beta^0(t) \in \tilde{\beta}(t)$ for any $t \in T$ and $f \circ \beta^0 \notin F$. Since F satisfies Bayesian monotonicity, there exist $i \in N$, $t_i \in T_i^*$, and a function $y^0 : T_{-i} \to \Delta(A)$ such that

$$U_i(y^0 \circ \beta^0_{-i}|t_i) > U_i(f \circ \beta^0|t_i),$$

and for all $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^0|\tilde{t}_i).$$

Recall that we have assumed that there exists a collection of pure deceptions $\{\beta^\lambda\}$ such that $\tilde{\beta} = \bigcup_\lambda \beta^\lambda$. In particular, we can assume without loss of generality that $\{\beta^\lambda\}$ is non-redundant, i.e., it never be the case that there exist $\lambda, \lambda' \in \Lambda$ with $\lambda \neq \lambda'$ such that $\beta^\lambda(\tilde{t}) = \beta^{\lambda'}(\tilde{t})$ for some $\tilde{t} \in T^*$. For each $t' = (t_i', t_{-i}') \in T$, define

$$\tilde{y}(t_i', t_{-i}') \equiv \begin{cases} y^0(t_{-i}') & \text{if } t_i' = \beta_i^0(t_i) \text{ and} \\ & t_{-i}' = \beta^0_{-i}(\tilde{t}_{-i}) \text{ for some } \tilde{t}_{-i} \\ f(t_i', t_{-i}') & \text{otherwise.} \end{cases}$$

For each $t_{-i}' \in T_{-i}$, we next define

$$y^*(t_{-i}') \equiv \sum_{t_i} \psi_i(t_i'|t_i)\tilde{y}(t_i', t_{-i}').$$

Due to the construction of y^* and Bayesian monotonicity for β^0, we obtain the following:

$$U_i(y^* \circ \psi_{-i}|t_i) = U_i(\tilde{y} \circ \psi|t_i) > U_i(f \circ \psi|t_i).$$

Recall that F satisfies incentive compatibility. This is where the argument of SV does not go through without incentive compatibility. Then, once again, due to the construction of y^* and Bayesian monotonicity for β^0, we obtain the following: for any $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^*|\tilde{t}_i).$$

This shows that mixed Bayesian monotonicity holds. ■

A.2. Proof of Lemma 1

Fix $i \in N$, $t_i \in T_i^*$, and $\phi_i, \tilde{\phi}_i \in \Delta(T_{-i})$. Fix $f \in F$ as well. We choose the constant γ so that

$$\gamma \sum_{t_{-i}, t_{-i}'} \tilde{\phi}_i(t_{-i}')\phi_i(t_{-i}) = 1.$$

For each $t_{-i}, t_{-i}' \in T_{-i}$, define

$$\psi_i(t_{-i}', t_{-i}) = \gamma \tilde{\phi}_i(t_{-i}')\phi_i(t_{-i}).$$

Since F satisfies NWR, there exist $y_i'[f; t_i, \psi_i], y_i[f; t_i, \psi_i] \in Y_i[f]$ such that

$$\sum_{t_{-i}, t_{-i}'} \psi_i(t_{-i}', t_{-i})u_i(y_i'[f; t_i, \psi_i](t_{-i}'); (t_i, t_{-i})) > \sum_{t_{-i}, t_{-i}'} \psi_i(t_{-i}', t_{-i})u_i(y_i[f; t_i, \psi_i](t_{-i}'); (t_i, t_{-i})).$$

Define

$$\alpha_i[t_i, \phi_i] = \sum_{t_{-i}'} \tilde{\phi}_i(t_{-i}')y_i[f; t_i, \psi_i](t_{-i}') \text{and} \alpha_i'[t_i, \phi_i] = \sum_{t_{-i}'} \tilde{\phi}_i(t_{-i}')y_i'[f; t_i, \psi_i](t_{-i}').$$

Then, by construction of $\alpha_i[t_i, \phi_i]$ and $\alpha_i'[t_i, \phi_i]$, we rewrite the above inequality:

$$\sum_{t_{-i}} \phi_i(t_{-i})u_i(\alpha_i'[t_i, \phi_i]; (t_i, t_{-i})) > \sum_{t_{-i}} \phi_i(t_{-i})u_i(\alpha_i[t_i, \phi_i]; (t_i, t_{-i})).$$

This completes the proof. ■

A.3. Proof of Theorem 2

Following the argument in Section 7, we denote by $\{t_i^\ell\}_{\ell=1}^\infty$ its countable support of $\Delta(T_i)$. We denote its countable support of $\Delta(T_{-i} \times T_{-i})$ by $\{\psi_i^k\}_{k=1}^\infty$. If an SCS F satisfies the intersection property, for all $f, f' \in F$ with $f \not\approx f'$ and each $i \in \{1, 2\}$, we define the uniform SCF $\bar{y}_i[f, f']$ as follows: there exist $\delta, \eta \in (0, 1)$ such that

$$\bar{y}_i[f, f'] \equiv (1-\delta)(1-\eta)\sum_{\ell=1}^\infty \eta^{\ell-1}\sum_{k=1}^\infty \delta^{k-1}y_i[f, f'; t_i^\ell, \psi_i^k].$$

Then, we define

$$\bar{y}[f, f'] = \frac{1}{2}\bar{y}_1[f, f'] + \frac{1}{2}\bar{y}_2[f, f'].$$

we use the uniform SCF $\bar{y}[f, f']$ in the canonical mechanism we propose below.

We prove this by constructing an implementing mechanism $\Gamma = (M, g)$. Each agent i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4, m_i^5)$ where $m_i^1 \in T_i^*$, $m_i^2 \in F$, $m_i^3 \in \mathbb{N} = \{1, 2, \ldots\}$, $m_i^4 = \{m_i^4[f]\}_{f \in F}$ where $m_i^4[f] \in Y_i[f]$ for each $f \in F$, and $m_i^5 \in \Delta^*(A)$. Since the type space T, the SCS F, each $Y_i[f]$, and $\Delta^*(A)$ are all finite, each message space M_i is made countable.

The outcome $g(m)$ is determined by the following rules:

Rule I: If there exist $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for all $j \in N$, then $g(m) = f(m^1)$.

Rule II: If $m_k^3 = 1$ for each $k \in N$ and there exist $f^i, f^j \in F$ with $f^i \not\approx f^j$ such that $m_i^2 = f^i$ and $m_j^2 = f^j$, then $g(m) = \bar{y}[f^i, f^j]$ where $\bar{y}[f^i, f^j]$ is the uniform SCF defined previously.

Rule III: If there exist $i \in N$ and $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for $j \neq i$ and $m_i^3 > 1$, then

$$g(m) = \begin{cases} m_i^4[f](m_{-i}^1) & \text{with probabiltiy } m_i^3/(m_i^3+1) \\ \bar{y}_i[f](m_{-i}^1) & \text{with probability } 1/(m_i^3+1) \end{cases}$$

where $\bar{y}_i[f]$ is the uniform SCF defined in Section 7.

Rule IV: In all other cases,

$$g(m)=\begin{cases} m_1^5 & \text{with probability } m_1^3/n(m_1^3+1)\\ m_2^5 & \text{with probability } m_2^3/n(m_2^3+1)\\ \vdots & \vdots \\ m_n^5 & \text{with probability } m_n^3/n(m_n^3+1)\\ \bar{\alpha} & \text{with the remaining probability}\end{cases}$$

where $\bar{\alpha}$ is the uniform lottery over A, as defined in Section 7.

We complete the proof by a series of claims below.

Claim I. *For any equilibrium $\sigma \in BNE^{\Gamma(T)}$, $m \in M$, and $t \in T^*$: $m \in supp(\sigma(\cdot|t)) \Rightarrow m_j^3 = 1$ for each $j \in N$.*

Proof of Claim I. This is essentially the same as the proof of Claim 1 of Theorem 1. Hence, we skip the proof. ■

Claim II. *For any equilibrium $\sigma \in BNE^{\Gamma(T)}$, there exists $f \in F$ such that $m_j^2 = f$ for each $j \in N$, $t_j \in T_j^*$, $m_j \in supp(\sigma_j(\cdot|t_j))$.*

Proof of Claim II. By the previous claim, we know that $m_j^3 = 1$ for each $j \in N$, $t_j \in T_j^*$, and $m_j \in \text{supp}(\sigma_j(\cdot|t_j))$. Suppose by way of contradiction that there exist $t \in T^*$ and $m \in \text{supp}(\sigma(\cdot|t))$ such that $m_i^2 = f^i \not\approx f^j = m_j^2$. We focus on agent i of type t_i. We fix such m_i and consider the messages of agent $j \neq i$ as follows: for each $t_j' \in T_j$,

$$M_j^*(t_j') = \left\{m_j |\ \exists f^j \in F \text{ s.t. } m_j^2 \approx f^j \text{ and } m_j^3 = 1 \text{ and } m_j^1 = t_j'\right\}.$$

For each $t_j' \in T_j$, if

$$\sum_{t_j} \pi_i(t_i)[t_j] \sum_{m_{-i}\in M_j^*(t_j')} \sigma_j(m_j|t_j) > 0,$$

then, we can set

$$\psi_i(t_j', t_j) \equiv \gamma \cdot \pi_i(t_i)[t_j] \sum_{m_j\in M_j^*(t_j')} \sigma_j(m_j|t_j)$$

for any $t_j \in T_j$ where the constant γ is chosen so that $\sum_{t_j,t_j'} \psi_i(t_j', t_j) = 1$. Since F satisfies the intersection property, there exists $\hat{m}_i^4[f^j] \in Y_i[f^j]$ such that

$$\sum_{t_j,t_j'} \psi_i(t_j', t_j) u_i(\hat{m}_i^4[f^j](t_j'); t_i, t_j) > \sum_{t_j,t_j'} \psi_i(t_j', t_j) u_i(\bar{y}[f^i, f^j](t_j'); t_i, t_j).$$

Define $\hat{m}_i$ to be the same as m_i except that $m_i^4[f^j]$ is replaced by $\hat{m}_i^4[f^j]$ defined above and $\hat{m}_i^3 > 1$ sufficiently large. Then, the above inequality allows us to conclude the following:

$$\sum_{t_j} \pi_i(t_i)[t_j] \sum_{m_j} \sigma_j(m_j|t_j) u_i(g(\hat{m}_i, m_j); t_i, t_j) > \sum_{t_j} \pi_i(t_i)[t_j] \sum_{m_j} \sigma_j(m_j|t_j) u_i(g(m_i, m_j); t_i, t_j)$$

This implies that $U_i(g \circ (\hat{m}_i, \sigma_j)|t_i) > U_i(g \circ (m_i, \sigma_j)|t_i)$. Thus, $\hat{m}_i$ would be an even better response for type t_i to σ_j than m_i. This contradicts our hypothesis that $\sigma \in BNE^{\Gamma(T)}$. ■

Claim III. *For each $f \in F$, there exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for all $j \in N$ and $t_j \in T_j$:*

$$m_j \in supp(\sigma_j(\cdot|t_j)) \Rightarrow m_j = (t_j, f, 1, m_j^4, m_j^5)$$

for some $m_j^4 \in M_j^4$ and $m_j^5 \in M_j^5$.

Proof of Claim III. Fix $f \in F$. For each $j \in N$, define a function $\sigma_j : T_j \to M_j$ satisfying the following properties: for each $t_j \in T_j^*$ and $m_j \in \text{supp}(\sigma_j(\cdot|t_j))$,

1. $m_j^1 = t_j$;
2. $m_j^2 = f$; and
3. $m_j^3 = 1$.

Fix $i \in N$ and $t_i \in T_i$. Then, it only remains to show that $\sigma_i(t_i)$ is a best response of type t_i to σ_{-i}. Observe first that m_i^5 is irrelevant for the resulting outcome given σ_{-i}. Let us denote by $\hat{m}_i$ a deviation strategy of type t_i. Assume that $\hat{m}_i$ has the property that $\hat{m}_i^2 = 1$ and $\hat{m}_i^2 = f' \not\approx f$ so that it induces Rule II. Then, the lottery $\bar{y}[f', f]$ is chosen so that type t_i is worse off with $\hat{m}_i$ than $\sigma_i(t_i)$. Suppose that $\hat{m}_i$ has the property that $\hat{m}_i^3 > 1$ so that it induces Rule III. Then, there is always positive probability that $\bar{y}_i[f]$ is chosen so that type t_i is worse off with $\hat{m}_i$ than $\sigma_i(t_i)$. Therefore, the only profitable deviation type t_i can possibly have is to change $\sigma_i^1(t_i)$ into $\hat{m}_i^1$ but change neither m_i^2 nor m_i^3. However, since the other agent $j \neq i$ is truthful by our hypothesis and f satisfies incentive compatibility, this cannot be profitable. Thus, $\sigma_i(t_i)$ is a best response of type t_i to σ_{-i}. ■

Claim IV. *For each $f \in F$ and deception β, if there exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for each $j \in N$, $t_j \in T_j$, and $m_j \in supp(\sigma_j(\cdot|t_j))$,*

$m_j = (t_j', f, 1, m_j^4, m_j^5)$ for some (m_j^4, m_j^5) and $t_j' \in \beta_j(t_j)$,

then, $f \circ \beta \in F$.

Proof of Claim IV. This is the same as the proof of Claim 4 of Theorem 1. Hence, we skip the proof. ■

All four claims we have established above imply that (1) for any $f \in F$, there exists $\sigma \in BNE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for any $\sigma \in BNE^{\Gamma(T)}$, we have that $g \circ \sigma \in F$. Thus, we complete the proof of the theorem. ■

A.4. Proof of Theorem 3

First, since every SCF satisfies closure, we no longer need this condition in the statement of the theorem. Second, the intersection property is vacuously satisfied for SCFs because it never be the case that two agents disagree on the SCF to be implemented. Recall also that every SCF satisfies the convex range property. So, given incentive compatibility, by Proposition 3, we can exploit the equivalence between mixed Bayesian monotonicity and Bayesian monotonicity. Thus, in the statement of the theorem, we replace mixed Bayesian monotonicity with just Bayesian monotonicity. We prove this by constructing an implementing mechanism $\Gamma = (M, g)$. Each agent i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4)$ where $m_i^1 \in T_i^*$ which is finite, $m_i^2 \in \mathbb{N} = \{1, 2, \ldots\}$, the set of positive integers, $m_i^3 \in Y_i[f]$ which is finite, and $m_i^4 \in \Delta^*(A)$ which is a finite subset of $\Delta(A)$. Therefore, each M_i can be made countable.

The outcome $g(m)$ is determined by the following rules:

Rule A: If $m_i^2 = 1$ for all $i \in N$, then $g(m) = f(m^1)$.

Rule B: If there exists $i \in N$ such that $m_j^2 = 1$ for all $j \neq i$ and $m_i^2 > 1$, then

$$g(m) = \begin{cases} m_i^3(m_{-i}^1) & \text{with probabiltiy } m_i^2/(m_i^2+1)\\ \bar{y}_i[f](m_{-i}^1) & \text{with probability } 1/(m_i^2+1)\end{cases}$$

where $\bar{y}_i[f]$ is the uniform SCF as defined in Section 7.

Rule C: In all other cases,

$$g(m)=\begin{cases} m_1^4 & \text{with probability } m_1^2/n(m_1^2+1)\\ m_2^4 & \text{with probability } m_2^2/n(m_2^2+1)\\ \vdots & \vdots \\ m_n^4 & \text{with probability } m_n^2/n(m_n^2+1)\\ \bar{\alpha} & \text{with the remaining probability}\end{cases}$$

where $\bar{\alpha}$ is the uniform lottery as defined in Section 7.

The proof of this theorem is essentially the same as that of Theorem 1 for SCSs. As in the proof of Theorem 1, the proof consists of a series of claims.

Claim A. *For any equilibrium $\sigma \in BNE^{\Gamma(T)}$, $m \in M$, and $t \in T^*$: $m \in supp\,(\sigma(\cdot|t)) \Rightarrow m_i^2 = 1$ for each $i \in N$.*

Claim B. *There exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for all $i \in N$ and $t_i \in T_i$, $\sigma_i(\cdot|t_i)$ assigns probability one on $m_i = (t_i, 1, m_i^3, m_i^4)$ for some $m_i^3 \in M_i^3$ and $m_i^4 \in M_i^4$.*

Claim C. *For each deception β, if there exists an equilibrium $\sigma \in BNE^{\Gamma(T)}$ such that for each $t \in T^*$, $m \in supp\,(\sigma(\cdot|t))$, and $j \in N$,*

$m_j = (t_j', 1, m_j^3, m_j^4)$for some (m_j^3, m_j^4) and $t_j' \in \beta_j(t_j)$,

then, $f \circ \beta \approx f$.

Claim A corresponds to Claim 1; Claim B corresponds to Claim 3; and Claim C corresponds to Claim 4 in Theorem 1. Hence, we skip all the proofs for Claims A–C. All three claims we have established above imply that (1) there exists $\sigma \in BNE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for any $\sigma \in BNE^{\Gamma(T)}$, we have $g \circ \sigma \approx f$. Thus, we complete the proof of the theorem. ∎

A.5. Proof of Proposition 5

($\Leftarrow$) Suppose that F satisfies Bayesian monotonicity. Fix $f \in F$. Assume that there exist $t, t' \in T^*$ such that $f(t') \in F(t)$. Define β as a single-valued deception with the following property: for any $i \in N$ and $\tilde{t}_i \in T_i$,

$$\beta_i(\tilde{t}_i)=\begin{cases} t_i' & \text{if } \tilde{t}_i = t_i\\ \tilde{t}_i & \text{otherwise.}\end{cases}$$

By construction of β above and Bayesian monotonicity of F, we have the following: there exist $i \in N, t_i \in T_i^*$, and a function $y^* : T_{-i} \to \Delta(A)$ such that

$$U_i(y^* \circ \beta_{-i}|t_i) > U_i(f \circ \beta|t_i),$$

while for all $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^*|\tilde{t}_i).$$

Define $\alpha = y^*(\beta_{-i}(t_{-i})) = y^*(t_{-i}') \in \Delta(A)$. Note that $f(\beta(t)) = f(t')$. Since we focus on a complete information environment, by Bayesian monotonicity, we obtain the following:

$$u_i(\alpha; t) > u_i(f(t'); t) \text{and} u_i(f(t'); t') \geq u_i(\alpha; t').$$

Thus, F also satisfies Maskin monotonicity.

($\Rightarrow$) Suppose that F satisfies Maskin monotonicity. Fix $f \in F$ and β as a single-valued deception such that $f \circ \beta \notin F$. Since we assume that $f \circ \beta \notin F$, there must exist $t, t' \in T^*$ such that $f(t') \notin F(t)$ and $t' = \beta(t)$. Fix such t and t'. Since F satisfies Maskin monotonicity, there exist $i \in N$ and $\alpha \in \Delta(A)$ such that

$$u_i(\alpha; t) > u_i(f(t'); t) \text{and} u_i(f(t'); t') \geq u_i(\alpha; t').$$

Define $y^* : T_{-i} \to \Delta(A)$ as follows: for any $\tilde{t}_{-i}$,

$$y^*(\tilde{t}_{-i})=\begin{cases} \alpha & \text{if } \tilde{t}_{-i} = t_{-i}'\\ f(t_i', \tilde{t}_{-i}) & \text{otherwise.}\end{cases}$$

Since we focus on a complete information environment, by construction of y^* and Maskin monotonicity of F, we have

$$U_i(y^* \circ \beta_{-i}|t_i) > U_i(f \circ \beta|t_i).$$

Since F satisfies incentive compatibility and we focus on a complete information environment, by construction of y^* and Maskin monotonicity of F, we have the following: for any $\tilde{t}_i \in T_i^*$,

$$U_i(f|\tilde{t}_i) \geq U_i(y^*|\tilde{t}_i).$$

Thus, F satisfies Bayesian monotonicity. ∎

A.6. Proof of Theorem 4

Recall the definition of $Y_i[f]$:

$$Y_i[f] \equiv \left\{y_i : T_{-i} \to \Delta^*(A) | U_i(f|\tilde{t}_i) \geq U_i(y_i|\tilde{t}_i)\ \forall \tilde{t}_i \in T_i^*\right\},$$

where $\Delta^*(A)$ is a finite subset of $\Delta(A)$. By NWA, $Y_i[f]$ is always nonempty. We prove this by constructing an implementing mechanism $\Gamma = (M, g)$. Each agent i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4, m_i^5) \in M_i$ where $m_i^1 \in T_i$, $m_i^2 \in F$, $m_i^3 \in \mathbb{N} = \{1, 2, \ldots\}$, $m_i^4 = \{m_i^4[f]\}_{f \in F}$ where $m_i^4[f] \in Y_i[f]$ for each $f \in F$, and $m_i^5 \in \Delta^*(A)$. Since the type space T, the SCS F, each $Y_i[f]$, and $\Delta^*(A)$ are all finite, each M_i is made countable.

Since T is finite, $\Delta(T)$ has a countable support. Thus, we denote by $\{t^\ell\}_{\ell=1}^\infty$ its countable support of $\Delta(T)$. Fix $f \in F$. Since F satisfies NWA, we can define a uniform lottery $\bar{y}_i[f]$ as follows:

$$\bar{y}_i[f] \equiv \frac{(1-\delta)}{2} \sum_{\ell=1}^{\infty} \delta^{\ell-1} \left\{f(t^\ell) + y_i[f](t^\ell)\right\}$$

for some $\delta \in (0, 1)$. We fix some $\bar{f} \in F$ and define another uniform lottery $\bar{\alpha}$ as follows:

$$\bar{\alpha} \equiv \frac{(1-\delta)}{2n} \sum_{i \in N} \sum_{\ell=1}^{\infty} \delta^{\ell-1} \left\{\bar{f}(t^\ell) + y_i[\bar{f}](t^\ell)\right\}.$$

These $\{\{\bar{y}_i[f]\}_{f \in F}\}_{i \in N}$ and $\bar{\alpha}$ will be used in the canonical mechanism.

The outcome $g(m)$ is determined by the following rules:

Rule i: If there exists $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for all $j \in N$, then $g(m) = f(m^1)$.

Rule ii: If there exist $i \in N$ and $f \in F$ such that $m_j^2 \approx f$ and $m_j^3 = 1$ for all $j \neq i$ and $m_i^2 \not\approx f$ or $m_i^3 > 1$, then

$$g(m)=\begin{cases} m_i^4[f](m_{-i}^1) & \text{with probabiltiy } m_i^3/(m_i^3+1)\\ \bar{y}_i[f] & \text{with probability } 1/(m_i^3+1)\end{cases}$$

where $\bar{y}_i[f]$ is the uniform lottery over A we defined previously using NWA.

Rule iii: In all other cases,

$$g(m)=\begin{cases} m_1^5 & \text{with probability } m_1^3/n(m_1^3+1)\\ m_2^5 & \text{with probability } m_2^3/n(m_2^3+1)\\ \vdots & \vdots \\ m_n^5 & \text{with probability } m_n^3/n(m_n^3+1)\\ \bar{\alpha} & \text{with the remaining probability}\end{cases}$$

where $\bar{\alpha}$ is the uniform lottery over A we defined previously using NWA. Note that closure becomes vacuous under a complete information type space. When there are at least three agents, by Proposition 6, we can find an SCS $\hat{F} \approx F$ such that $\hat{F}$ satisfies incentive compatibility. In what follows, we focus on the SCS $\hat{F}$ and will prove that $\hat{F}$ is Nash implementable by the canonical mechanism proposed above. After we observe that any Bayesian Nash equilibrium reduces to a Nash equilibrium under complete information, the proof of the theorem is essentially the same as that of Theorem 1 but the main difference from Theorem 1 is that one can weaken NWR into NWA and mixed Bayesian monotonicity into Maskin monotonicity, respectively.

The proof consists of a series of claims.

Claim i. *For any equilibrium $\sigma \in NE^{\Gamma(T)}$, $m \in M$, and $t \in T^*$: $m \in supp(\sigma(\cdot|t)) \Rightarrow m_i^3 = 1$ for each $i \in N$.*

Proof of Claim i. Since we consider a complete information type space, the following fact is commonly certain among all the agents: there exists a collection of $\{\varphi_i\}_{i\in N}$ for which each $\varphi_i : T_i^* \to T_{-i}$ is a function such that for each $t \in T^*$, $i, j \in N$, we have $(t_i, \varphi_i(t_i)) = (t_j, \varphi_j(t_j)) = t$.

Fix a Nash equilibrium $\sigma \in NE^{\Gamma(T)}$. We focus on type t_i of agent i. Suppose by way of contradiction that $\text{supp}(\sigma_i(\cdot|t_i))$ contains a message m_i such that $m_i^3 > 1$. We fix such m_i and partition the messages of all agents but i as follows: for each $t'_{-i} \in T_{-i}$,

$$M^*_{-i}(t'_{-i}) = \Big\{m_{-i} |\ \exists f \in F \text{ s.t. } m_j^2 \approx f \text{ and } m_j^3 = 1 \text{ for all } j \neq i \text{ and } m^1_{-i} = t'_{-i}\Big\},$$

and

$$\hat{M}_{-i} = \{m_{-i} |\ m_j^3 > 1 \text{ for some } j \in N\backslash\{i\} \text{ or } \exists j, k \in N\backslash\{i\} \text{ s.t. } m_j^2 \not\approx m_k^2\}.$$

If

$$\sum_{m_{-i}\in\hat{M}_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) > 0,$$

then, since F satisfies NWA, we know that there exists $\hat{m}_i^5 \in \Delta(A)$ such that

$$u_i(\hat{m}_i^5; t_i, \varphi_i(t_i)) > u_i(\bar{\alpha}; t_i, \varphi_i(t_i)).$$

For each $t'_{-i} \in T_{-i}$, if

$$\sum_{m_{-i}\in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) > 0,$$

then, we can set

$$\psi_i(t'_{-i}) \equiv \gamma \sum_{m_{-i}\in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|\varphi_i(t_i))$$

where the constant γ is chosen so that $\sum_{t'_{-i}} \psi_i(t'_{-i}) = 1$. Since F satisfies NWA, there exists $\hat{m}_i^4[f] \in Y_i[f]$ such that

$$\sum_{t'_{-i}} \psi_i(t'_{-i}) u_i(\hat{m}_i^4[f](t'_{-i}); t_i, \varphi_i(t_i)) > u_i(\bar{y}_i[f]; t_i, \varphi_i(t_i)).$$

Define $\hat{m}_i$ to be the same as m_i except that $m_i^4[f]$ is replaced by $\hat{m}_i^4[f]$ and m_i^5 is replaced by $\hat{m}_i^5$; and $\hat{m}_i^3 = m_i^3 + 1$. Then, the above inequality allows us to conclude the following:

$$\sum_{m_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) u_i(g(\hat{m}_i, m_{-i}); t_i, \varphi_i(t_i)) > \sum_{m_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) u_i(g(m_i, m_{-i}); t_i, \varphi_i(t_i)).$$

Thus, $\hat{m}_i$ would be an even better response to σ_{-i} than m_i. This contradicts our hypothesis that $\sigma \in NE^{\Gamma(T)}$. ∎

Claim ii. *For any equilibrium $\sigma \in NE^{\Gamma(T)}$, there exists $f \in \hat{F}$ such that $m_j^2 = f$ for each $j \in N$, $t_j \in T_j^*$, and $m_j \in supp(\sigma_j(\cdot|t_j))$.*

Proof of Claim ii. Fix an equilibrium $\sigma \in NE^{\Gamma(T)}$. By the previous claim, for any $j \in N$, $t \in T$, and $m \in \text{supp}(\sigma(\cdot|t))$, we have $m_j^3 = 1$. Suppose on the contrary that there exist $t \in T^*$, $m \in \text{supp}(\sigma(t))$, and $i, k \in N$ with $i \neq k$ such that $m_i^2 \not\approx m_k^2$. Then, we fix such m_i and only need to consider the following two cases: either

Case 1: there exist $j, k \in N\backslash\{i\}$ with $j \neq k$ such that $m_i^2 \not\approx m_j^2$, $m_j^2 \not\approx m_k^2$, and $m_k^2 \not\approx m_i^2$; or

Case 2: there exists $f \in F$ such that $m_j^2 \approx f$ for all $j \neq i$.

By our hypothesis, either Cases 1 or 2 occur with probability one. We focus on agent i of type t_i. We first assume that Case 1 applies. We consider the following messages of all agents but i:

$$\hat{M}_{-i} = \{m_{-i} |\ \exists j, k \in N\backslash\{i\} \text{ s.t. } m_i^2 \not\approx m_j^2, m_j^2 \not\approx m_k^2, \text{ and } m_k^2 \not\approx m_i^2\}.$$

If

$$\sum_{m_{-i}\in\hat{M}_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) > 0,$$

then, since F satisfies NWA, there exists $\hat{m}_i^5 \in \Delta(A)$ such that

$$u_i(\hat{m}_i^5; t_i, \varphi_i(t_i)) > u_i(\bar{\alpha}; t_i, \varphi_i(t_i)).$$

We next assume that Case 2 applies. We consider the following messages of all agents but i: for each $t'_{-i} \in T_{-i}$,

$$M^*_{-i}(t'_{-i}) = \Big\{m_{-i} |\ \exists f \in F \text{s.t. } m_j^2 \approx f\ \forall j \neq i \text{ and } m^1_{-i} = t'_{-i}\Big\}.$$

For each $t'_{-i} \in T_{-i}$, if

$$\sum_{m_{-i}\in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) > 0,$$

then, we can set

$$\psi_i(t'_{-i}) \equiv \gamma \sum_{m_{-i}\in M^*_{-i}(t'_{-i})} \sigma_{-i}(m_{-i}|\varphi_i(t_i))$$

where the constant γ is chosen so that $\sum_{t'_{-i}} \psi_i(t'_{-i}) = 1$. Since F satisfies NWA, there exists $\hat{m}_i^4[f] \in Y_i[f]$ such that

$$\sum_{t'_{-i}} \psi_i(t'_{-i}) u_i(\hat{m}_i^4[f](t'_{-i}); t_i, \varphi_i(t_i)) > u_i(\bar{y}_i[f]; t_i, \varphi_i(t_i)).$$

Define $\hat{m}_i$ to be the same as m_i except that $m_i^4[f]$ is replaced by $\hat{m}_i^4[f]$; m_i^5 by $\hat{m}_i^5$ defined above and $\hat{m}_i^3 = m_i^3 + 1$. Then, we conclude the following:

$$\sum_{m_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) u_i(g(\hat{m}_i, m_{-i}); t_i, \varphi_i(t_i)) > \sum_{m_{-i}} \sigma_{-i}(m_{-i}|\varphi_i(t_i)) u_i(g(m_i, m_{-i}); t_i, \varphi_i(t_i)).$$

Thus, $\hat{m}_i$ would be an even better response to σ_{-i} than m_i. This contradicts our hypothesis that $\sigma \in NE^{\Gamma(T)}$. ∎

Claim iii. *For each $f \in \hat{F}$, there exists an equilibrium $\sigma \in NE^{\Gamma(T)}$ such that for all $j \in N$ and $t_j \in T_j$:*

$$m_j \in supp(\sigma_j(\cdot|t_j)) \Rightarrow m_j = (t_j, f, 1, m_j^4, m_j^5)$$

for some $m_j^4 \in M_j^4$ and $m_j^5 \in M_j^5$.

Proof of Claim iii. Fix $f \in F$. For each $j \in N$, define $\sigma_j : T_j \to M_j$ satisfying the following properties: for each $t_j \in T_j^*$ and $m_j \in \text{supp}(\sigma_j(\cdot|t_j))$,

1. $m_j^1 = t_j$;
2. $m_j^2 = f$; and
3. $m_j^3 = 1$.

Fix $i \in N$ and $t_i \in T_i$. It only remains to show that $\sigma_i(t_i)$ is a best response to σ_{-i}. Observe that m_i^5 is irrelevant for the resulting outcome given σ_{-i}. If type t_i deviates from $\sigma_i(t_i)$ to some $\hat{m}_i$ where $\hat{m}_i^2 \not\approx f$ or $\hat{m}_i^3 > 1$, he can induce Rule ii. Then, there is always positive probability that $\bar{y}_i[f]$ is chosen and since F satisfies NWA, type t_i will be worse off than playing $\sigma_i(t_i)$. Thus, the only profitable deviation type t_i can possibly have is to change m_i^1 but change neither m_i^2 nor m_i^3. However, since all agents are truthful by our hypothesis and f satisfies incentive compatibility, this cannot be profitable. Therefore, $\sigma_i(t_i)$ is a best response of t_i to σ_{-i}. ■

Claim iv. *For each SCF $f \in \hat{F}$ and deception β, if there exists a Nash equilibrium $\sigma \in NE^{\Gamma(T)}$ such that for each $t \in T^*$, $m \in$ supp $(\sigma(\cdot|t))$, and $j \in N$,*

$m_j = (t_j', f, 1, m_j^4, m_j^5)$ *for some* (m_j^4, m_j^5) *and* $t_j' \in \beta_j(t_j)$,

then, $f \circ \beta \in \hat{F}$.

Proof of Claim iv. Note that the convex range property becomes a vacuous constraint under a complete information type space. By Propositions 3 and 6, we know that $\hat{F}$ satisfies mixed Bayesian monotonicity. Therefore, this is the same as the proof of Claim 4 of Theorem 1. ■

All four claims we have established above imply that (1) for any $f \in \hat{F}$, there exists $\sigma \in NE^{\Gamma(T)}$ such that $g \circ \sigma \approx f$; and (2) for any $\sigma \in NE^{\Gamma(T)}$, we have that $g \circ \sigma \in \hat{F}$. Thus, we complete the proof of the theorem. ■

A.7. Extension to a more general setup

I extend the analysis of this paper to a more general environment with incomplete information. Assume that T_i is a *Polish* space T_i associated with its Borel σ-algebra $\mathscr{T}_i$. I endow T_{-i} and T with the product Borel σ-algebras $\mathscr{T}_{-i}$ and $\mathscr{T}$, respectively. Note that T_{-i} and T are also Polish spaces. Let $\Delta(T_{-i})$ denote the set of probability distributions on measurable space $(T_{-i}, \mathscr{T}_{-i})$ endowed with the weak-* topology. Each agent i's system of "interim" beliefs is expressed as a $\mathscr{T}_i$-measurable function $\pi_i : T_i \to \Delta(T_{-i})$. Then, I call $(T_i, \mathscr{T}_i, \pi_i)_{i\in N}$ a *type space*. Let A denote the set of pure outcomes associated with its σ-algebra $\mathscr{A}$ containing all singleton sets. Let $\Delta(A)$ be the set of probability distributions over measurable space $(A, \mathscr{A})$. Agent i's state dependent von Neumann–Morgenstern utility function is denoted $u_i : \Delta(A) \times T \to \mathbb{R}$, which is assumed to be a $\mathscr{A} \times \mathscr{T}$-measurable function. I can now define an *environment* as $\mathcal{E} = (A, \mathscr{A}, \{u_i, T_i, \mathscr{T}_i, \pi_i\}_{i\in N})$.

A subset of T is called an event if it is $\mathscr{T}$-measurable.[19] An event $E = E_1 \times \cdots \times E_n \subseteq T$ is said to be *belief-closed* if, for each $i \in N$ and $t_i \in E_i$, we have $\pi_i[t_i](E_{-i}) = 1$. I assume that the planner only cares about the belief-closed subset of the type space $(T_i^*, \mathscr{T}_i^*)_{i\in N}$ where $T_i^* \subseteq T_i$ and $\mathscr{T}_i^*$ is its relative σ-algebra for every $i \in N$.

A (stochastic) *social choice function* (SCF) is a $\mathscr{T}$-measurable function $f : T \to \Delta(A)$. Let $\mathbb{F}$ be the collection of all $\mathscr{T}$-measurable SCFs. A *social choice set* (SCS) F is defined as a nonempty subset of $\mathbb{F}$. Two SCSs F and H are said to be *equivalent* ($F \approx H$) if there exists a bijection $\xi : F \to H$ such that $\sup\{|f(\mathcal{A}|t) - h(\mathcal{A}|t)| : t \in T^*, \mathcal{A} \in \mathscr{A}\} = 0$ for every $f \in F$ and every $h \in H$ satisfying $h = \xi(f)$. This means that the two SCSs "coincide" for every $t \in T^*$.

A *mechanism* (or *game form*) $\Gamma = ((M_i, \mathscr{M}_i)_{i\in N}, g)$ describes a nonempty message space M_i for each agent i, equipped with a σ-algebra $\mathscr{M}_i$ and an $\mathscr{M}$-measurable outcome function $g : M \to \Delta(A)$, where $M = \times_{i\in N} M_i$ is associated with product σ-algebra $\mathscr{M}$.

The interim expected utility of agent i of type t_i that pretends to be of type t_i' in the direct-revelation mechanism associated with an SCF f, provided all other agents are truthful is defined as:

$$U_i(f; t_i'|t_i) \equiv \int_{\mathscr{T}_{-i}} u_i(f(t_i', t_{-i}); (t_i, t_{-i}))\pi_i[t_i](dt_{-i})$$

Denote $U_i(f|t_i) = U_i(f; t_i|t_i)$.

A.7.1. Mixed Bayesian implementation

Given a mechanism $\Gamma = (M, \mathscr{M}, g)$, let $\Gamma(T)$ denote an incomplete information game associated with a type space $(T_i, \mathscr{T}_i, \pi_i)_{i\in N}$. Let $\sigma_i : T_i \to \Delta(M_i)$ denote a $\mathscr{T}_i$-measurable mixed strategy for agent i and Σ_i his set of *mixed* strategies, where $\Delta(M_i)$ denotes the set of probability measures over $(M_i, \mathscr{M}_i)$ endowed with the weak-* topology. Let $\sigma_i(\cdot|t_i) \in \Delta(M_i)$ denote the probability measure over $(M_i, \mathscr{M}_i)$ conditional upon t_i. Besides, I denote by $\sigma(\cdot|t) = \prod_{j\in N} \sigma_j(\cdot|t_j) \in \prod_{j\in N} \Delta(M_j)$ the product measure over $(M, \mathscr{M})$ conditional upon t. I assume that $g \circ \sigma$ is a $\mathscr{T} \times \mathscr{M}$-measurable function and $g \circ \sigma \in \mathbb{F}$ for every $\sigma \in \Sigma$. With abuse of notation, I let

$$U_i(g \circ (\sigma_i', \sigma_{-i})|t_i)$$
$$\equiv \int_{\mathscr{T}_{-i}} \int_{\mathscr{M}_{-i}} \int_{\mathscr{M}_i} u_i(g(m_i, m_{-i}); \times (t_i, t_{-i}))\sigma_i'(dm_i|t_i)\sigma_{-i}(dm_{-i}|t_{-i})\pi_i[t_i](dt_{-i}).$$

A.7.2. Necessity for mixed Bayesian implementation

This paper discusses three necessary conditions: (1) incentive compatibility; (2) closure; and (3) mixed Bayesian monotonicity. When dealing with a more general setup, one needs no modification for incentive compatibility. The only modification I need for defining closure in a more general case is the measurability requirement for events. More specifically, a subset of T is said to be an event if it is $\mathscr{T}$-measurable. Finally, the only modification one needs for mixed Bayesian monotonicity is the requirement that each deception $\beta_i : T_i \to 2^{T_i}\backslash\{\emptyset\}$ is $\mathscr{T}_i$-measurable and each mapping $\psi_k : T_k \to \Delta(T_k)$ is $\mathscr{T}_k$-measurable.

A.7.3. Sufficiency for mixed Bayesian implementation

In this section, I discuss how one can extend Theorem 1 (sufficiency for mixed Bayesian implementation) to a more general setup. For each SCF $f \in \mathbb{F}$, define

$$Y_i[f] \equiv \left\{ y_i : T_{-i} \to \Delta(A) \left| \begin{array}{l} y_i \text{ is } \mathscr{T}_{-i}\text{-measurable and} \\ U_i(f|\tilde{t}_i) \geq U_i(y_i|\tilde{t}_i) \ \forall \tilde{t}_i \in T_i^* \end{array} \right. \right\}.$$

The set $Y_i[f]$ is associated with its Borel σ-algebra $\mathscr{Y}_i[f]$.

Since T_i is a Polish space, $\Delta(T_i)$ can also be made Polish. I denote by $\{t_i^\ell\}_{\ell=1}^\infty$ its countable dense subset of $\Delta(T_i)$. Similarly,

[19] Since $\mathscr{T}$ is the product measure, any event constitutes a product set.

since $T_{-i} \times T_{-i}$ is a Polish space, $\Delta(T_{-i} \times T_{-i})$ can also be made Polish. So, I denote by $\{\psi_i^k\}_{k=1}^{\infty}$ its countable dense subset of $\Delta(T_{-i} \times T_{-i})$. Since F satisfies NWR, for each $f \in F$ and $i \in N$, I define the uniform SCF $\bar{y}_i[f]$ as follows: there exist $\delta, \eta \in (0, 1)$ such that

$$\bar{y}_i[f] \equiv \frac{(1-\delta)(1-\eta)}{2} \sum_{\ell=1}^{\infty} \eta^{\ell-1} \times \sum_{k=1}^{\infty} \delta^{k-1} \left\{ y_i'[f; t_i^\ell, \psi_i^k] + y_i[f; t_i^\ell, \psi_i^k] \right\}.$$

Recall that this uniform SCF $\bar{y}_i[f]$ is used in the canonical mechanism for Theorem 1. Note also that $\{t_i^\ell\}_{\ell=1}^{\infty}$ is a dense subset of $\Delta(T_i)$ and $\{\psi_i^k\}_{k=1}^{\infty}$ is a dense subset of $\Delta(T_{-i} \times T_{-i})$, respectively. As expected utilities are continuous in both $\Delta(T_i)$ and $\Delta(T_{-i} \times T_{-i})$, NWR together with the uniform SCF $\bar{y}_i[f]$ plays exactly the same role in the proof of Theorem 1 as if the type space is countable.

Similarly, for each $i \in N$, I define the uniform lottery $\bar{\alpha}_i \in \Delta(A)$ as follows: there exist $\delta, \eta \in (0, 1)$ such that

$$\bar{\alpha}_i \equiv \frac{(1-\delta)(1-\eta)}{2} \sum_{\ell=1}^{\infty} \eta^{\ell-1} \sum_{k=1}^{\infty} \delta^{k-1} \left\{ \alpha_i'[t_i^\ell, \phi_i^k] + \alpha_i[t_i^\ell, \phi_i^k] \right\}.$$

Finally, I define

$$\bar{\alpha} \equiv \frac{1}{n} \sum_{i \in N} \bar{\alpha}_i.$$

Recall that this uniform lottery $\bar{\alpha}$ is used in the canonical mechanism of Theorem 1. Note also that $\{t_i^\ell\}_{\ell=1}^{\infty}$ is a dense subset of $\Delta(T_i)$ and $\{\psi_i^k\}_{k=1}^{\infty}$ is a dense subset of $\Delta(T_{-i} \times T_{-i})$, respectively. Once again, as expected utilities are continuous in both $\Delta(T_i)$ and $\Delta(T_{-i} \times T_{-i})$, NWR (more precisely, Lemma 1) together with the uniform lottery $\bar{\alpha}$ plays exactly the same role in the proof of Theorem 1 as if the type space is countable.

I state the extension of Theorem 1 to a more general setup.

Theorem 5. *Suppose there are at least three agents ($n \geq 3$). If an SCS F satisfies incentive compatibility, mixed Bayesian monotonicity, closure, and NWR, then it is mixed Bayesian implementable.*

Proof. We use the same mechanism proposed in the proof of Theorem 1. In the proposed mechanism $\Gamma = (M, \mathscr{M}, g)$, each agent i sends a message $m_i = (m_i^1, m_i^2, m_i^3, m_i^4, m_i^5) \in M_i$ where $m_i^1 \in T_i^*$, $m_i^2 \in F$, $m_i^3 \in \mathbb{N} = \{1, 2, \ldots\}$, $m_i^4 = \{m_i^4[f]\}_{f \in F}$ where $m_i^4[f] \in Y_i[f]$ for each $f \in F$, and $m_i^5 \in \Delta(A)$. The only modification we need is to impose the measurability requirement over the message space. Let $\mathscr{M}_i = \mathscr{T}_i \times \mathscr{F} \times 2^{\mathbb{N}} \times \prod_{f \in F} \mathscr{Y}_i[f] \times \mathscr{A}$ be its associated σ-algebra.

The rest of the proof is completed by appropriately adapting that of Theorem 1 to the current setup. ■

It is easy to see that similar extensions can be established for Theorems 2–4 as well.

References

Abreu, D., Matsushima, H., 1992. Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information. Mimeo, Princeton University.

Abreu, D., Matsushima, H., 1994. Exact implementation. J. Econom. Theory 64, 1–19.

Artemov, G., Kunimoto, T., Serrano, R., 2013. Robust virtual implementation: Toward a reinterpretation of the Wilson doctrine. J. Econom. Theory 148, 424–447.

Aumann, R., 1987. Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55, 1–18.

Bergemann, D., Morris, S., 2009. Robust virtual implementation. Theor. Econ. 4, 45–88.

Bergemann, D., Morris, S., 2011. Robust implementation in general mechanisms. Games Econom. Behav. 71, 261–281.

Bergemann, D., Morris, S., Tercieux, O., 2011. Rationalizable implementation. J. Econom. Theory 146, 1253–1274.

Brandenburger, A., Dekel, E., 1993. Hierarchies of beliefs and common knowledge. J. Econom. Theory 59, 189–198.

Cabrales, A., Serrano, R., 2011. Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. Games Econom. Behav. 73, 360–374.

Chakravorti, B., 1992. Efficiency and mechanisms with no regret. Internat. Econom. Rev. 33, 45–59.

Chen, Y-C., Kunimoto, T., Sun, Y., 2016. Implementation with Transfers. Mimeo.

Duggan, J., 1997. Virtual Bayesian implementation. Econometrica 65, 1175–1199.

Dutta, B., Sen, A., 1994. Two-person Bayesian implementation. Econ. Des. 1, 41–54.

Jackson, M., 1991. Bayesian implementation. Econometrica 59, 461–477.

Jackson, M., 1992. Implementation in undominated strategies: A look at bounded mechanisms. Rev. Econom. Stud. 59, 757–775.

Jackson, M., 2001. A crash course in implementation theory. Soc. Choice Welf. 18, 655–708.

Kunimoto, T., Serrano, R., 2011. A new necessary condition for implementation in iteratively undominated strategies. J. Econom. Theory 146, 2583–2595.

Lipman, B., 2003. Finite order implications of common priors. Econometrica 71, 1255–1267.

Maskin, E., 1999. Nash equilibrium and welfare optimality. Rev. Econ. Stud. 66, 23–38.

Matsushima, H., 1993. Bayesian monotonicity with side payments. J. Econom. Theory 59, 107–121.

Mookherjee, D., Reichelstein, S., 1990. Implementation via augumented revelation mechanism. Rev. Econom. Stud. 57, 453–476.

Muller, E., Satterthwaite, M.A., 1977. The equivalence of strong positive association and strategy-proofness. J. Econom. Theory 14, 412?418.

Oury, M., Tercieux, O., 2012. Continuous implementation. Econometrica 80, 1605–1637.

Palfrey, T., Srivastava, S., 1987. On Bayesian implementable allocations. Rev. Econom. Stud. 54, 193–208.

Palfrey, T., Srivastava, S., 1989a. Implementation with incomplete information in exchagne economies. Econometrica 57, 115–134.

Palfrey, T., Srivastava, S., 1989b. Mechanism design with incomplete information: A solution to the implementation problem. J. Political Econ. 97, 668–691.

Palfrey, T., Srivastava, S., 1993. Bayesian Implementation. Harwood Academic Publishers.

Postlewaite, A., Schmeidler, D., 1986. Implementation in differential information economies. J. Econom. Theory 39, 14–33.

Saijo, T., 1987. On constant maskin monotonic social choice functions. J. Econom. Theory 42, 382–386.

Serrano, R., 2004. The theory of implementation of social choice rules. SIAM Rev. 46, 377–414.

Serrano, R., Vohra, R., 2005. A characterization of virtual Bayesian implementation. Games Econom. Behav. 50, 312–331.

Serrano, R., Vohra, R., 2010. Multiplicity of mixed equilibria in mechanisms: A unified approach to exact and approximate implementation. J. Math. Econom. 46, 775–785.