Closure under interim utility equivalence implies two-agent Bayesian implementation[*]

Ville Korpela[†] and Michele Lombardi[‡]

December 22, 2019

Abstract

We study the relationship between the two-agent implementation problem and the concept of interim efficiency of Holmström and Myerson (1983) in Bayesian environments with private values and independent types. We present a general property, called *closure under interim utility equivalence*, and show it is sufficient for the implementation of social choice functions. This condition, when combined with another property, called *interim inseparability*, is also sufficient for the implementation of essentially single-valued social choice sets. The characterization results are then examined in a variety of environments.

1

1 Introduction

The theory of implementation investigates the goals a designer can attain when they depend on the information possessed by various agents. The Nash implementation problem of the designer entails devising a mechanism in which agents' incentives lead to Nash equilibrium allocations associated with the goals set by the designer (Maskin, 1999). When such a mechanism exists, the designer's goals are (fully) Nash implementable. This paper studies the two-agent implementation problem in a setting with incomplete information, in which agents possess exclusive information. In such environments, the Nash implementation has a natural analog, referred to as Bayesian implementation. In Bayesian environments, a social choice function (SCF) f is a function from states to allocations, whereas a social choice set (SCS) F is a collection of SCFs.

With the exception of Mookherjee and Reichelstein (1990) and Dutta and Sen (1994), all studies on Bayesian implementation assume there are more than two agents (e.g., see Palfrey and Srivastava, 1989b; Jackson, 1991). Whereas Mookherjee and Reichelstein (1990) propose a general procedure on how to modify an incentive compatible revelation mechanism to Bayesian implement an SCS, Dutta and Sen (1994) provide a sufficient and almost necessary condition for two-agent Bayesian implementation. Conversely, this paper is concerned with the relationship between the two-agent Bayesian implementation problem and the concept of interim efficiency of Holmström and Myerson (1983).

When agents are asymmetrically informed, an (interim) incentive compatibility condition is necessary for Bayesian implementation. An SCF f is incentive compatible when truth telling is a Bayesian equilibrium of the direct mechanism associated with f. While incentive compatibility is sufficient to ensure the existence of desirable equilibria, more restrictions are required on SCFs to rule out undesirable equilibria.

The notion of interim incentive efficiency of Holmström and Myerson (1983) inspires the main restriction imposed on SCSs in this paper. That is, an SCF f is interim efficient if there is no other SCF $\hat{f}$ that would make every type of every agent as well off as f does, and would make at least one type strictly better off; that is, the SCF is interim efficient if it is welfare undominated. In other words, interim efficiency is a natural analog of Pareto efficiency when agents have private information. An SCF

is *interim incentive efficient* if it is incentive compatible and interim efficient.

This paper proposes a sufficient condition, called *closure under interim utility equivalence*. An SCS F satisfies this condition if each SCF f in F is interim incentive efficient, and whenever an incentive compatible SCF f' makes every possible type of every agent at least as well off as does f in F (in the interim stage), then f' must be an element of F. When F is an SCF, closure requires $f' = F$. Additionally, when F is an SCF, closure implies that the interim utility vector of F is unique for every agent under the set of all incentive compatible allocation rules.

Hitherto, no study uses the property of interim efficiency for exploring the Bayesian implementation of SCSs. While a few papers on mechanism design explore the property of interim efficiency but are limited either to a few applications or a specific environment.[1]

We show that every SCF satisfying closure is Bayesian implementable in environments with private values and independent types. Interestingly, this characterization result provides an understanding of what can be implemented (e.g., in nontransferable utility settings) when the designer considers not only incentive compatibility but also full implementation. Our findings are different from those of previous studies. Palfrey and Srivastava (1989a, 1991) show that, in private value models, the multiple equilibrium problem can be avoided if agents do not play weakly dominated strategies in equilibrium. Ollár and Penta (2017) address the same issue and also consider common knowledge assumptions by studying full implementation via transfer schemes under general restrictions on agents' beliefs.

When an SCF f of interest is interim incentive efficient but does not satisfy closure, it may be natural to turn to the implementation of a second best and try to implement an SCS that consists of allocation rules that are interim incentive efficient and welfare equivalent to f for every possible type of every agent. We provide a characterization result for this second best. In addition to closure, this result relies on another auxiliary condition, called *interim inseparability*. Generally, F is interim

---

[1] For instance, Gresik (1996) and Wilson (1985) focus on bilateral trade applications, particularly double auctions. Laussel and Palfrey (2003) and Ledyard and Palfrey (1994, 1999a, 1999b, 2002) study the applications to public good mechanisms. Ledyard and Palfrey (2007) investigate interim efficiency in private value environments, in which agents' utility functions are both quasi-linear in a private good and linear in a one-dimensional private value type parameter.

inseparable if f' makes every possible type of every agent as well off as does f when these allocation rules are both in F. This result then yields a characterization that applies to essentially single-valued SCSs (at the interim stage).

The remainder of the paper is organized as follows. Section 2 presents the basic notation and model. Section 3 presents the characterization result for SCFs, whereas Section 4 focuses on SCSs. Section 5 concludes the paper. The Appendix presents the proofs not included in the main text.

2   Notation and definitions

The set of *agents* is denoted by $I = \{1, 2\}$ and the set of (pure) *decisions* by D. The set of lotteries over D is denoted by $\Delta(D)$. Each agent i privately observes a parameter θ_i that determines her/his preferences. We refer to θ_i as agent i's *type*. The set of possible types for agent i is assumed to be finite and is denoted by Θ_i. A *state* is any pair of types $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \equiv \Theta$. For each agent i, we use Θ_{-i} to express the set of possible types of agent $j \neq i$ and, thus, a state is a profile (θ_i, θ_{-i}) where $\theta_i \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$. The same notational convention applies to any profile of objects.

An SCF or *allocation rule* is a function $x : \Theta \to \Delta(D)$ from states to lotteries over D. The set of all allocation rules is denoted by $X = \{x \mid x : \Theta \to \Delta(D)\}$. Further, an SCS F is any (non-empty) subset of X.

Each agent i is assumed to be an expected utility maximizer, whose (Bernoulli) utility function is $u_i(x, \theta)$ in state $\theta \in \Theta$. We focus on the case in which each agent's preference depends only on her/his type, known as *private values*. In this case, agent i's utility function, $u_i(x, \theta)$, can be denoted by $u_i(d, \theta_i)$ in state $\theta \in \Theta$. The probability density over the possible realizations of $\theta \in \Theta$ is denoted by $q(\cdot)$. We assume that $q(\cdot)$ has *full support*, that is, $q(\theta) > 0$ for all $\theta \in \Theta$.[2] For each type θ_i of agent i, the conditional probability of $\theta_{-i} \in \Theta_{-i}$ is the posterior belief of type θ_i and is denoted by $q(\theta_{-i}|\theta_i)$. We assume types to be stochastically independent. That is,

---

[2] An alternative informational setting is that of nonexclusive information (see, e.g., Postlewaite and Schmeidler, 1986; Palfrey and Srivastava, 1986). In this setting, each agents' own information is redundant if other individuals pool their information.

$q(\theta) = q_1(\theta_1) \times q_2(\theta_2)$, where, for each agent $i \in I$, q_i is a probability distribution over Θ_i.

Given allocation rule $x \in X$, the (*interim*) *expected utility* of agent i of type θ_i is defined by:

$$U_i^x(\theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} q(\theta_{-i}|\theta_i) \cdot v_i(x(\theta_{-i}, \theta_i), \theta_i),$$

where $v_i(x(\theta_{-i}, \theta_i), \theta_i)$ is the expected utility of lottery $x(\theta_{-i}, \theta_i)$, computed using the utility function $u_i(x, \theta_i)$.

An *environment* is a list $\langle I, D, q, (u_i, \Theta_i)_{i \in I} \rangle$ in which each element is assumed to be common knowledge among agents. Agent i knows her/his own type θ_i but is unsure about θ_{-i}.

A *mechanism* is a pair $\Gamma = (M, g)$, where $M = M_1 \times M_2$ is the *message space* and $g : M \to \Delta(D)$ is the *outcome function*. A *(pure) strategy* for agent i is a function $s_i : \Theta_i \to M_i$ and the set of all strategies for agent i is $S_i = \{s_i : \Theta_i \to M_i\}$. A *strategy profile* is a list $s \equiv (s_1, s_2)$ of strategies, one for each agent. The space of all strategy profiles is $S = S_1 \times S_2$. Given strategy profile s, we denote by $g(s)$ the outcome generated by s, where the outcome in state θ is denoted by $g(s(\theta))$. For exposition, we only focus below on pure strategies; most concepts and results can be extended to cover the case of mixed strategies (as discussed in Section 5).

Given a mechanism Γ, a *Bayesian (Nash) equilibrium* of Γ is a strategy profile s, so that for each agent $i \in I$ and each type $\theta_i \in \Theta_i$, it holds that:

$$U_i^{g(s)}(\theta_i) \geq U_i^{g(s_i', s_{-i})}(\theta_i) \text{ for all } s_i' \in S_i.$$

Let us denote the set of all Bayesian equilibria of mechanism Γ by $\mathcal{B}(\Gamma)$.

A mechanism Γ *implements* F in Bayesian equilibria if

$$\{g(s) \,|\, s \in \mathcal{B}(\Gamma)\} = F.$$

If there exists a mechanism Γ that implements F in Bayesian equilibria, F is said to be Bayesian implementable.

A necessary condition for Bayesian implementation is incentive compatibility, defined as follows.

Definition 1. F is (*Bayesian*) *incentive compatible* if, for all $f \in F$, all $i \in I$, and

5

all $\theta_i \in \Theta_i$,

$$U_i^f (\theta_i) \geq U_i^f \left(\hat{\theta}_i, \theta_i \right), \text{ for all } \hat{\theta}_i \in \Theta_i,$$

where $U_i^f \left(\hat{\theta}_i, \theta_i \right)$ is the (interim) expected utility of type θ_i when the agent reports $\hat{\theta}_i$, that is:

$$U_i^f \left(\hat{\theta}_i, \theta_i \right) = \sum_{\theta_{-i} \in \Theta_{-i}} q(\theta_{-i} \mid \theta_i) \cdot v_i \left(f(\theta_{-i}, \hat{\theta}_i), \theta_i \right).$$

We denote $U_i^f (\theta_i) \equiv U_i^f (\theta_i, \theta_i)$.

The proof of the necessity of this condition for implementation in Bayesian equilibria has been provided, for instance, by Dasgupta et al. (1979) and Jackson (1991). The intuition behind the condition is straightforward: an SCF is incentive compatible if truthful revelation is the best response for each agent whenever the other agents also reveal their true types.[3]

3 Implementation of SCFs

Since agents' types are determined at the interim stage, our notion of efficiency is that of interim efficiency, based on Holmoström and Myerson (1983). The key is identifying the appropriate notion of interim efficiency, which we call *closure under interim utility equivalence*. Before stating this condition, we need the following additional notation. Let X^{IC} be the set of all incentive compatible allocation rules, that is:

$$X^{IC} \equiv \{x \in X \mid x \text{ is incentive compatible}\}.$$

An SCF $\hat{f}$ *welfare dominates* f if $\hat{f}$ makes the expected utility of every possible type of every agent at least as large as f does and makes the expected utility of some types of some agents strictly larger. That is, $\hat{f}$ welfare dominates f if $U_i^{\hat{f}} (\theta_i) \geq U_i^f (\theta_i)$ for all $\theta_i \in \Theta_i$ and all $i \in I$, and $U_i^{\hat{f}} (\theta_i) > U_i^f (\theta_i)$ for some $\theta_i \in \Theta_i$ and some $i \in I$. An SCF f is *interim incentive efficient* if f is incentive compatible and no allocation rule in X^{IC} welfare dominates it. That is, f is interim incentive efficient if $f \in X^{IC}$ and there is no $\hat{f} \in X^{IC}$ such that $U_i^{\hat{f}} (\theta_i) \geq U_i^f (\theta_i)$ for all $\theta_i \in \Theta_i$ and all $i \in I$ and

---

[3]See d'Aspremont and Gérard-Varet (1979, 1982) for a more indepth discussion on incentive compatibility.

$U_i^{\hat{f}}(\theta_i) > U_i^f(\theta_i)$ for some $\theta_i \in \Theta_i$ and some $i \in I$. Therefore, we posit the following definition.

Definition 2. F satisfies *closure under interim utility equivalence*, or simply *closure*, if every $f \in F$ is interim incentive efficient and, for all $f \in F$ and all $x \in X^{IC}$, it holds that:

$$U_i^x(\theta_i) = U_i^f(\theta_i) \text{ for all } i \in I \text{ and all } \theta_i \in \Theta_i \implies x \in F.$$

This condition is composed of two parts. The first requires no incentive compatible allocation rule can welfare dominate $f \in F$. The second requires that, if an incentive compatible allocation rule $x \in X^{IC}$ and an SCF $f \in F$ are equally good for every possible type of every agent, then x is also in F. When F is an SCF, closure requires $x = F$. Additionally, when F is an SCF, closure implies that the interim utility vector of F is unique for every agent under the set of all incentive compatible allocation rules. Therefore, our first result is that every SCF satisfying closure is Bayesian implementable.

Theorem 1. If f satisfies closure under interim utility equivalence, then it is Bayesian implementable.

Proof. See the Appendix. ∎

Remark 1. From the proof of Theorem 1, it is clear it holds for any finite number of agents.

The proof of Theorem 1 consists of augmenting the direct revelation mechanism associated with f and specifying outcomes so that f is fully implemented in Bayesian equilibria if it satisfies closure. While a detailed proof of the theorem is given in the Appendix, we briefly describe it here. When agent i announces type θ_i in the direct revelation mechanism associated with f, s/he expects to face a lottery whose outcome is $f(\theta_{-i}, \theta_i)$ with probability $q_i(\theta_{-i}|\theta_i)$. By using a random device, we can construct a lottery $\ell_i(f, \theta_i)$ yielding outcome $f(\theta_{-i}, \theta_i)$ with probability $q_i(\theta_{-i}|\theta_i)$. Since values

7

are independent and private, agent i will be indifferent between lottery $\ell_i(f, \theta_i)$ and the one obtained by participating in the direct revelation mechanism—provided that the other agent is truthful. The devised mechanism uses other components of agents' message spaces, in addition to type announcements, to guarantee each agent i can always obtain $\ell_i(f, \theta_i)$, regardless what the other agent reports. Due to this feature, each Bayesian equilibrium produces an interim expected utility at least as great as that produced by f. The result follows by the assumption that f satisfies closure.

More importantly, Theorem 1 shows what can be implemented (e.g., in non-transferable utility settings) when the designer considers not only incentive compatibility but also full implementation. This is different from the solution provided by Palfrey and Srivastava (1989a, 1991), who show that, in private value models the multiple equilibrium problem can be avoided if agents do not play weakly dominated strategies in equilibrium.

In the following, we briefly discuss the implications of the above result.

Example 1 (Information design). Let $\langle I, D, q, (u_i, \Theta_i)_{i \in I} \rangle$ be an environment such that the ex-post sum of utilities

$$v_1(x, \theta_1) + v_2(x, \theta_2)$$

has a unique maximizer in all states $\theta \in \Theta$. Under this assumption, the maximizer is a pure decision.[4] We can define f by setting:

$$f(\theta) = \arg\max_{x \in D}[u_1(x, \theta_1) + u_2(x, \theta_2)],$$

for each $\theta \in \Theta$.

As we assume the existence of a unique maximizer for the ex-post sum of utilities, f satisfies closure if it is incentive compatible. However, the incentive compatibility of f depends on whether the designer can affect agents' prior beliefs so that f becomes incentive compatible. If the designer succeeds in this information design exercise, f is Bayesian implementable according to Theorem 1.

---

[4]Indeed, if lottery $x \in \Delta(D)$ maximizes the ex-post sum of utilities, all decisions in the support of x must also maximize this sum. Since we require the maximizer to be unique, x is a degenerated lottery.

To show that the designer can succeed in this task, let us further specify Example 1 by considering two cases—a voting environment and a public good environment.

Example 2 (Veto rule). A committee of two agents has to select a job applicant from a set of three candidates, $D = \{a, b, c\}$. In each state $\theta \in \Theta$, each committee member has a strict preference ranking over D. Therefore, each state $\theta \in \Theta$ can be identified by the pair of strict preferences $(\succ_1^\theta, \succ_2^\theta)$. We assume Θ includes all possible strict preference profiles, that is, for any pair of strict preferences $(\succ_1, \succ_2)$, there exists $\theta \in \Theta$ so that $(\succ_1, \succ_2) = (\succ_1^\theta, \succ_2^\theta)$ (unrestricted domain).

To complete the model, we need to fix agents' utility functions. Assume that, for any state θ, agent 1's utility function assigns utility of 3 to the preferred candidate, 1 to the middle candidate, and 0 to the least preferred candidate. Agent 2's utility function assigns utility of 1 to the preferred candidate, $\frac{1}{2}$ to the middle candidate, and -3 to the least preferred candidate.

The hiring process proceeds as follows. First, agent 2 vetoes the worst candidate $d \in D$. Second, agent 1 selects one candidate from $D - \{d\}$. Let $r_i^k(\theta)$ be the candidate that committee member i ranks kth in state θ. If both committee members act truthfully, the hiring process implements f defined by:

$$f(\theta) = \underset{x \in \{r_2^1(\theta), r_2^2(\theta)\}}{\arg\max} \ u_1(x, \theta_1),$$

for each $\theta \in \Theta$. However, this rule is manipulable ex-post.[5]

Suppose that each state is equally likely. Under this assumption, f is incentive compatible. Agent 1 has no reason to lie. Moreover, regardless which candidate agent 2 vetoes, agent 1 would select each of the remaining candidates with probability $\frac{1}{2}$. This implies lying is not profitable (on average) for agent 2 either.

Since f is incentive compatible, to show f is implementable according to Theorem 1, it suffices to verify that $f(\theta)$ uniquely maximizes the ex-post sum of utilities in each state $\theta \in \Theta$. As such, it suffices to note that the candidate maximizing the

---

[5]Indeed, there are cases in which agent 2 would not veto her/his worst candidate if s/he knew agent 1's ranking. As such, let θ be so that the ranking of agent 1 is $a \succ_1^\theta b \succ_1^\theta c$ and that of agent 2 is $b \succ_2^\theta a \succ_2^\theta c$. When the hiring procedure is followed, agent 2 vetoes c, meaning that candidate a would be hired. However, if agent 2 knew agent 1's ranking, s/he vetoes a because agent 2 would know that agent 1 would never select candidate c. In this case, agent 2's best candidate is hired.

ex-post sum in each state θ is the top candidate for agent 1 if s/he is not the worst candidate for agent 2; otherwise, s/he is the second best candidate for agent 1.

Example 3 (Facility location). Let $D = \{1, \ldots, k\} \subseteq \mathbb{Z}_+$ be the set of locations to build a public facility. There are two groups, with types $\Theta_1 = \Theta_2 = \{1, \ldots, k\}$. Type $\theta_i = h$ of group i means h is the most preferred location. We assume that preferences are single-peaked, that is, for each group $i \in \{1, 2\}$ and each location $d \in D$,

$$u_i(d, \theta_i) = -|\theta_i - d| \text{ for all } \theta_i \in \Theta_i.$$

The goal of the designer is to build the public facility in the best location of the group nearest to location 1. This goal is obviously strategy-proof and, thus, incentive compatible. However, there are other locations that yield exactly the same ex-post sum of utilities. This is because the ex-post sum $u_1(x, \theta_1) + u_2(x, \theta_2)$ is constant for all locations d between peaks θ_1 and θ_2. Therefore, the goal of the designer does not necessarily satisfy closure, although it is incentive compatible.

As such, assume there exists a third group with known type $\theta_3 = 1$ (see Remark 1). The utility function of this group is $u_3(d, 1) = -|1 - d|$. After the introduction of the new group, the goal of the designer can be expressed as follows: for all $\theta \in \Theta \equiv \Theta_1 \times \Theta_2 \times \{\theta_3\}$,

$$f(\theta) = \text{median}\{\theta_1, \theta_2, \theta_3\}.$$

It is well-known this SCF is strategy-proof and, hence, incentive compatible.

To show f is Bayesian implementable, we only need to show $f(\theta)$ is the unique maximizer of the ex-post sum of utilities in each state $\theta \in \Theta$. To this end, we fix any $\theta \in \Theta$. Assume, to the contrary, that there exists another location $d \neq f(\theta)$ that maximizes the ex-post sum of utilities, that is:

$$\sum_{i=1}^{3} u_i(d, \theta_i) = \sum_{i=1}^{3} u_i(f(\theta), \theta_i). \tag{1}$$

We proceed according to whether $d < f(\theta)$ or not.

Assume $d < f(\theta)$. This implies $1 = \theta_3 \leq d < f(\theta)$. Without loss of generality, we assume $f(\theta) = \theta_2$. Then, it holds that $1 = \theta_3 \leq d < f(\theta) = \theta_2 \leq \theta_1$. Since agents have single-peaked preferences, it follows from (1) that

$$-(\theta_1 - d) - (\theta_2 - d) + (\theta_3 - d) = -(\theta_1 - \theta_2) + (\theta_3 - \theta_2),$$

10

which simplifies to $d = \theta_2$, which is a contradiction.

Assume $d > f(\theta)$. This implies $1 = \theta_3 \le f(\theta) < d$. Without loss of generality, we assume $f(\theta) = \theta_2$. Then, it holds that $1 = \theta_3 \le f(\theta) = \theta_2 < d$. We thus distinguish two cases.

Assume $\theta_1 \ge d$. Then, we have $1 = \theta_3 \le f(\theta) = \theta_2 < d \le \theta_1$. It follows from (1) that:

$$- (\theta_1 - d) + (\theta_2 - d) + (\theta_3 - d) = -(\theta_1 - \theta_2) + (\theta_3 - \theta_2),$$

which simplifies to $d = \theta_2$, which is a contradiction.

Assume $\theta_1 < d$. In this case, it holds that $1 = \theta_3 \le f(\theta) = \theta_2 \le \theta_1 < d$. It follows from (1) that:

$$(\theta_1 - d) + (\theta_2 - d) + (\theta_3 - d) = -(\theta_1 - \theta_2) + (\theta_3 - \theta_2),$$

which simplifies to $3d = 2\theta_1 + \theta_2$. Since $\theta_2 \le \theta_1 < d$, it follows that $3d > 2\theta_1 + \theta_2$, which is a contradiction. Thus, f satisfies closure.

Example 4 (Bilateral trading). Let us consider a bilateral trading problem in which agent 1 has an indivisible object that agent 2 wants to buy. The valuation of agent i is denoted by v_i, drawn from a set $V_i \subseteq \mathbb{R}_+$ according to the distribution F_i. Both agents are risk neutral. A *trading rule* or *selling mechanism* is a pair (p, t), where $p : V_1 \times V_2 \to [0, 1]^2$ is the *allocation rule* and $t : V_1 \times V_2 \to \mathbb{R}^2$ is the *payment rule*. If v_1 and v_2 are the valuations reported by agents, $p_i(v_1, v_2)$ is the probability the object is transferred to agent i, and $t_i(v_1, v_2)$ is the expected payment of agent i.

In the following, we show it is generically impossible to apply Theorem 1 to bilateral trading environments. Let ℓ be a lottery that yields $t \in \mathbb{R}_+$ monetary units with probability $\frac{1}{2}$ and $-t$ monetary units with probability $\frac{1}{2}$, and (p, t) any interim incentive efficient trading rule. We define a new trading mechanism (p, t') by setting $t'(v_1, v_2) = t(v_1, v_2) + \ell$ for each pair of valuations $(v_1, v_2) \in V_1 \times V_2$.[6] (p, t') is an incentive compatible trading mechanism. Since $\mathbb{E}(\ell) = 0$ by construction, it follows that every type of every agent is indifferent between (p, t') and (p, t). However, since $(p, t') \ne (p, t)$, the second part of closure is violated. In the next subsection, we propose a solution to this problem.

---

[6]$t(v_1, v_2) + \ell$ is a compound lottery, where the outcomes of $t(v_1, v_2)$ and ℓ are added.

4 Implementation of SCSs

Let us turn briefly to the problem of implementing an SCS, that is, of ensuring all SCFs $f \in F$ and only these are obtained as Bayesian equilibria. We consider SCSs because there exist cases in which Theorem 1 cannot be applied. To this end, assume that f is not fully implementable according to Theorem 1, although it is interim incentive efficient (see Example 4). In cases such as this one, it is natural to turn to the second best solution. One appealing resolution here is to enlarge the objective of implementation by fully implementing all allocation rules that are welfare equivalent to f. Therefore, besides closure, we require F to satisfy the following auxiliary condition.

Definition 3. F is *interim inseparable* if $U_i^f(\theta_i) = U_i^{\hat{f}}(\theta_i)$ holds for all $i \in I$, all $\theta_i \in \Theta_i$ and all $f, \hat{f} \in F$.

In other words, F is interim inseparable when agents wish to fully implement f and agree to fully implement any allocation rule $\hat{f}$ that makes every possible type of every agent as well off as f does (at the interim stage). The characterization result can be stated as follows.

Theorem 2. If F satisfies closure under interim utility equivalence and interim inseparability, then it is Bayesian implementable.

Proof. See the Appendix. ∎

Remark 2. It is obvious from the proof of Theorem 2 that it holds for any finite number of agents.

Let us briefly discuss two implications of Theorem 2.

Example 5 (Borda's rule). A committee of two agents has to select one candidate from $D = \{d_1, d_2, \ldots, d_k\}$. Let $k \geq 3$. At each state $\theta \in \Theta$, each committee member has a strict preference ranking over D. Therefore, each $\theta \in \Theta$ can be identified by a pair of strict preferences $(\succ_1^\theta, \succ_2^\theta)$. We assume that Θ includes all possible strict preference profiles, that is, for any pair of strict preferences $(\succ_1, \succ_2)$, there exists $\theta \in \Theta$ so that $(\succ_1, \succ_2) = (\succ_1^\theta, \succ_2^\theta)$ (unrestricted domain).

Suppose each state $\theta \in \Theta$ is equally likely. Furthermore, for any state $\theta \in \Theta$ and any agent i, assume that the utility representation of the ranking $\succ_i^\theta$ is a one-to-one assignment of numbers $1, 2, ..., k$ to the elements of D. In this setting, rank $r_i(d_\ell, \theta_i)$ of candidate $d_\ell \in D$ for type θ_i of agent i is given by:

$$r_i(d_\ell, \theta_i) = (k + 1) - u_i(d_\ell, \theta_i).$$

The set of all *Borda winners* in state $\theta \in \Theta$, denoted by $B(\theta)$, is defined by

$$B(\theta) = \arg \min_{d \in D} \left[r_1(d, \theta_1) + r_2(d, \theta_2) \right].$$

Nothing guarantees the Borda winner is a 'fair' outcome. As such, an outcome best for one agent and worst for the other can be a Borda winner.[7] Thus, let us define the set of 'fair Borda winners' by:

$$FB(\theta) = \arg \min_{d \in B(\theta)} \left[\max \left\{ r_1(d, \theta_1), r_2(d, \theta_2) \right\} \right],$$

for each $\theta \in \Theta$. The goal of the designer is to implement f that selects a uniform distribution, defined over the set of fair Borda winners, that is:

$$f(\theta) \sim U\left(FB(\theta) \right).$$

This SCF is incentive compatible. Intuitively, if an agent lies, then s/he will lose equally often as s/he wins, as every type of the other agent is equally likely. On average, lying does not pay off. Moreover, this SCF is on the interim efficiency frontier, as $f(\theta)$ maximizes the sum of utilities at each state $\theta \in \Theta$. However, there are other selections of Borda winners that yield exactly the same interim expected utilities for all types of committee members. Theorem 2 helps in situations such as this one guaranteeing that an SCS F that contains f and any other allocation rule $f' \in F$ so that (1) f' is incentive compatible, (2) $\operatorname{supp}(f'(\theta)) \subseteq B(\theta)$ for all $\theta \in \Theta$ (f' makes a selection of Borda winners) and (3) $U_i(f', \theta_i) = U_i(f, \theta_i)$ for both $i \in I$ and all $\theta_i \in \Theta_i$ (interim inseparability) is Bayesian implementable.[8] Although the

---

[7] In this case, $B(\theta) = D$.

[8] Note that $\operatorname{supp}(f'(\theta))$ consists of the pure decisions that have non-zero probability according to lottery $f'(\theta)$.

allocation rules in F are not necessarily fair ex-post, they are all fair on average or in expectation.[9]

Example 6 (Optimal selling mechanism). A seller s has one indivisible object to be sold to buyer 1 or buyer 2. The valuation of buyer i is denoted by v_i and drawn from a finite set $V_i \equiv \Theta_i \subseteq \mathbb{R}_+$ according to distribution F_i. Buyers valuations are private and independent. Both the seller and buyers are risk neutral.

A trading rule is a pair (p, t), where $p : V_1 \times V_2 \to [0, 1]^2$ is the *allocation rule* and $t : V_1 \times V_2 \to \mathbb{R}^2$ is the *payment rule*. For any pair of valuations $v = (v_1, v_2)$, we interpret $p_i(v)$ as the probability with which agent i obtains the object, and $t_i(v)$ is the expected payment of agent i. The goal of the seller is to find an individually rational and incentive compatible trading rule that maximizes expected revenue $\mathbb{E}[t_1(v) + t_2(v)]$. This problem has been solved by Skreta (2007) by generalizing the method of Myerson (1981). Let us denote the revenue maximizing trading rule by (p^*, t^*).[10]

Note (p^*, t^*) is not necessarily on the interim efficiency frontier of buyers. The reason is that each buyer would be better off by slightly decreasing her/his payment without reducing the welfare of the other buyer. To ensure we are at the interim efficiency frontier, we consider the seller as a player, that is $I = \{s, 1, 2\}$. In this way, if a buyer pays less, s/he decreases the welfare of the seller. Let us assume that the reservation price, v_s, of the seller is common knowledge among the agents. That is, let $\Theta_s = \{v_s\}$. Recall that Theorem 2 holds for any number of agents (see Remark 2). Additionally, if trading rule (p, t) yields an interim expected utility equal to that generated by the revenue maximizing trading rule (p^*, t^*), for each type of each agent, trading rule (p, t) is individually rational (at the interim stage) and expected revenue maximizing. Then, let us define F by:[11]

$$F = \left\{ (p, t) \mid (p, t) \text{ is IC and } U_i^{(p,t)}(\theta_i) = U_i^{(p^*,t^*)}(\theta_i) \text{ for all } \theta_i \in \Theta_i \text{ and all } i \in I \right\}.$$

F is Bayesian implementable according to Theorem 2 by construction. It is worth noting that all possible outcomes are expected revenue maximizing.

---

[9] If one committee member benefits in one state, then the other committee member must benefit in another, that is, not the same committee member benefits always.

[10] For details, we refer the reader to Skreta (2007).

[11] IC means incentive compatible.

Remark 3. In practice, we do not need to know F either in Example 5 or Example 6. To better understand this, let us focus on Example 6. Let us consider the revenue maximizing trading rule (p^*, t^*) of Skreta (2007). Consider the mechanism constructed in the proof of Theorem 1. Assume that the seller is an agent in I. In this case, any trading rule that arises as a Bayesian equilibrium of the mechanism yields an interim expected utility at least as high as that for (p^*, t^*) for every type of every agent in I. Note the interim expected utility of the seller cannot be higher than that for (p^*, t^*); otherwise, this trading rule is not a revenue maximizing one. Therefore, whatever is the equilibrium outcome of the mechanism, it will achieve revenue maximization, although we do not know the set F.

5   Concluding remarks

The focus of this paper is on Bayesian implementation in pure strategies. However, characterization results extend to the case of mixed strategies. For instance, assume the hypotheses of Theorem 1 are met and let us consider the mechanism constructed in its proof. First, note that a pure strategy equilibrium that coincides with f is still an equilibrium, although mixed strategies are allowed. Second, assume that a (non-degenerate) mixed strategy profile $\sigma = (\sigma_1, \sigma_2)$ is a Bayesian equilibrium. It follows that every message in the support of $\sigma_i(\theta_i)$ yields the same interim expected utility for every type θ_i of agent i. By the proposed mechanism, this interim expected utility is at least as high as that produced by the lottery $\ell_i(f, \theta_i)$. The proof goes through since f satisfies closure.

The assumption of private values with independent types is central to our analysis, allowing us to construct a mechanism where agent i of type θ_i can obtain the lottery $\ell_i(f, \theta_i)$ that makes her/him at least as well off (at the interim stage) as the lottery in the direct revelation mechanism, regardless what the other agent $-i$ is doing. Although the construction in the proof of Theorem 1 is not always feasible when values are common and/or types are correlated, a similar construction can be provided when the SCF can be *virtually simulated*. Generally, f can be virtually simulated if one can find, for each agent i, an "incentive compatible set of lotteries" that gives to agent i a level of expected utility that coincides with that s/he can obtain by

15

participating in the direct revelation mechanism associated with f under truthful behavior. More formally, for any agent i and any SCF f, let $\ell_i(f)$ denote a (non-empty) subset of $\Delta(D)$. We say that the set $\ell_i(f)$ *virtually simulates the SCF f* if, for every $\theta_i \in \Theta_i$, there is a lottery $\ell_i^f(\theta_i) \in \ell_i(f)$ so that (1) $U_i^f(\theta_i) = U_i^{\ell_i^f(\theta_i)}(\theta_i)$ and (2) $U_i^{\ell_i^f(\theta_i)}(\theta_i) \geq U_i^{\ell_i^f(\theta_i')}(\theta_i)$ for every $\theta_i' \in \Theta_i$, where for all $\bar{\theta}_i \in \Theta_i$:

$$U_i^{\ell_i^f(\bar{\theta}_i)}(\theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} q(\theta_{-i}|\theta_i) \cdot v_i\left(\ell_i^f(\bar{\theta}_i), (\theta_{-i}, \theta_i)\right), \text{ for all } \theta_i \in \Theta_i.$$

If, for each agent i, $\ell_i(f)$ virtually simulates f, we say that f is *virtually simulated by* $(\ell_1(f), \ell_2(f))$.

It can be shown that an SCF is Bayesian implementable in any environment if it satisfies closure and can be virtually simulated.[12] This follows from exactly the same arguments used for private values and independent types when lottery $\ell_i(f, \theta_i)$ is replaced with lottery $\ell_i^f(\theta_i)$. It will be interesting to know the classes of environments under which a given SCF can be virtually simulated. This topic will be addressed in future research.

Appendix

Notation: For any $x \in X$, $i \in I$, and $\theta_i \in \Theta_i$, we define lottery $\ell_i(x, \theta_i)$ as follows: the outcome is $x(\theta_{-i}, \theta_i)$ with probability $q(\theta_{-i}|\theta_i)$ for all $\theta_{-i} \in \Theta_{-i}$. In a private value environment with independent types, if x is incentive compatible, then agent i of type θ_i weakly prefers $\ell_i(x, \theta_i)$ to $\ell_i(x, \theta_i')$ for all $\theta_i' \in \Theta_i$. A deception for agent $i \in I$ is a mapping $\alpha^i : \Theta_i \rightarrow \Theta_i$. Let A^i be the set of all deceptions for agent i and $A = A^1 \times A^2$, with α as a typical profile of deceptions. For any $\alpha \in A$, any $x \in X$, and any θ, the lottery delivered by $(x \circ \alpha)$ is $(x \circ \alpha)(\theta)$.

Proof of Theorem 1

Assume that f satisfies closure. Consider $\Gamma = (M_1 \times M_2, g)$, where $M_i = \Theta_i \times \{0, 1\} \times \mathbb{Z}_+$, that is, each agent is asked to report a type, an element of the set $\{0, 1\}$ and a

---

[12]Note that any two lotteries, $\ell_i(f, \theta_i)$ and $\ell_{-i}(f, \theta_{-i})$, must have common elements in their support. At least $f(\theta_i, \theta_{-i})$ must be in the support of both. However, this does not need to hold for $\ell_i^f(\theta_i)$ and $\ell_{-i}^f(\theta_{-i})$, meaning there are more degrees of freedom in designing the mechanism.

non-negative integer z^i.[13] The outcome function $g : M_1 \times M_2 \to \Delta(D)$ is defined by the following rules:

Rule 1. If $m^i = (\theta_i, 0, z^i)$ for both $i \in I$, then $g(m) = f(\theta_1, \theta_2)$.

Rule 2. If $m^{-i} = (\theta_{-i}, 0, z^{-i})$ and $m^i = (\theta_i, 1, z^i)$, then $g(m) = \ell_i(f, \theta_i)$.

Rule 3. In all other cases, an integer game is played: identify the agent who announces the highest integer (if there is a tie, pick agent 1). Let agent i be the winner. Then, the lottery implemented is $g(m) = \ell_i(f, \theta_i)$.

Let us first show that f can be supported as a Bayesian equilibrium of the mechanism. Let $s = (s_1, s_2)$ be a strategy profile so that agent i of type $\theta_i \in \Theta_i$ announces $s_i(\theta_i) = (\theta_i, 0, z^i)$. If this profile is played, *Rule 1* applies, and $g(s) = f$. Note that agent i of type θ_i is indifferent between f and $\ell_i(f, \theta_i)$ by construction of the lottery $\ell_i(f, \theta_i)$. Agent i of type θ_i does not want to change only her/his type announcement by incentive compatibility—the new message profile would still fall under *Rule 1*. Assume that agent i of type θ_i changes the second component of her/his message into 1. In this case, *Rule 2* applies and produces the lottery $\ell_i(f, \alpha^i(\theta_i))$. Since f is incentive compatible, type θ_i weakly prefers $\ell_i(f, \theta_i)$ to $\ell_i(f, \alpha^i(\theta_i))$. Therefore, $\ell_i(f, \alpha^i(\theta_i))$ is no better than f for type θ_i of agent i. Since the choice of type θ_i and of agent i are arbitrary, we conclude that $s \in \mathcal{B}(\Gamma)$ and its outcome is f.

Conversely, take any $s \in \mathcal{B}(\Gamma)$. We show that $g(s) = f$ to conclude the proof. Fix any type θ_i of agent i. Assume that type θ_i deviates by changing $s_i(\theta_i)$ to $m^i = (\theta_i, 1, z^i)$. Therefore, the lottery selected by the outcome function is $\ell_i(f, \theta_i)$—provided that type θ_i adjusts z^i so that s/he becomes the winner of the integer game if *Rule 3* applies (as there are finitely many states). As $s \in \mathcal{B}(\Gamma)$, type θ_i weakly prefers the lottery obtained under message $s_i(\theta_i)$ to lottery $\ell_i(f, \theta_i)$.

Since the choice of type θ_i of agent i and of agent i are arbitrary, the lottery obtained under message $s_i(\theta_i)$ is preferred to lottery $\ell_i(f, \theta_i)$ for every possible type θ_i of every agent i; that is, $U_i^{g(s)}(\theta_i) \geq U_i^f(\theta_i)$ for all $i \in I$ and all $\theta_i \in \Theta_i$. Since f is interim incentive efficient, f cannot be welfare dominated by $g(s)$, and so $U_i^{g(s)}(\theta_i) =$

---

[13]$\mathbb{Z}_+$ is the set of non-negative integers.

$U_i^f(\theta_i)$ for all $i \in I$ and all $\theta_i \in \Theta_i$. By the definition of Bayesian equilibrium, the allocation rule $g(s) \in X^{IC}$. As both $g(s)$ and f are elements of X^{IC} and as f satisfies closure, $g(s) = f$, as we sought. ∎

Proof of Theorem 2

Assume that the hypotheses of Theorem 2 are met. Consider $\Gamma = (M_1 \times M_2, g)$, where $M_i = \Theta_i \times F \times \{0, 1\} \times \mathbb{Z}_+$, that is, each agent is asked to report a type, an element of F, an element of the set $\{0, 1\}$ and a non-negative integer z^i. The outcome function $g : M_1 \times M_2 \to \Delta(D)$ is defined by the following rules:

Rule 1. If $m^i = (\theta_i, f, 0, z^i)$ for both $i \in I$, then $g(m) = f(\theta_1, \theta_2)$.

Rule 2. For both $i \in I$, if $m^{-i} = (\theta_{-i}, f^{-i}, 0, z^{-i})$ and $m^i = (\theta_i, f^i, 1, z^i)$, then $g(m) = \ell_i(f^{-i}, \theta_i)$.

Rule 3. In all other cases, an integer game is played: identify the agent who announces the highest integer (if there is a tie, pick agent 1). Let agent i be the winner. Then, the implemented lottery is $g(m) = \ell_i(f^i, \theta_i)$.

Let us first show that each $f \in F$ can be supported as a Bayesian equilibrium of the mechanism. To this end, we fix any $f \in F$. Let $s = (s_1, s_2)$ be a strategy profile so that type $\theta_i \in \Theta_i$ announces $s_i(\theta_i) = (\theta_i, f, 0, z^i)$. If this profile is played, *Rule 1* applies, and $g(s) = f$. Moreover, type θ_i does not want to change the third component of her/his message into 1. In this case, *Rule 2* applies and produces the lottery $\ell_i(f, \alpha_i(\theta_i))$, where $\alpha^i(\theta_i)$ is her/his new type announcement. As f is incentive compatible by assumption and we are in an environment with private values and independent types, type θ_i weakly prefers $\ell_i(f, \theta_i)$ to $\ell_i(f, \alpha^i(\theta_i))$. Moreover, agent i of type θ_i is indifferent between f and $\ell_i(f, \theta_i)$ by construction of the lottery $\ell_i(f, \theta_i)$. This implies that $\ell_i(f, \alpha^i(\theta_i))$ is no better than f for any type θ_i of agent i. Therefore, the proposed strategy profile s is in $\mathcal{B}(\Gamma)$ and its outcome is f.

Conversely, take any $s \in \mathcal{B}(\Gamma)$. We show that $g(s) = f$ for some $f \in F$.

We fix any θ_i. Assume that type θ_i deviates by changing $s_i(\theta_i)$ into $m^i = (\theta_i, f, 1, z^i)$. Therefore, the lottery selected by g is either $\ell_i(f^{-i}, \theta_i)$ if *Rule 2* applies or $\ell_i(f, \theta_i)$ if *Rule 3* applies—provided that type θ_i adjusts z^i by which s/he becomes the winner of the integer game. Since $f^{-i}, f \in F$, $U_i^f(\theta_i) = U_i^{f^{-i}}(\theta_i)$ for all $i \in I$ and all $\theta_i \in \Theta_i$, by the interim inseparability of F, and so $\ell_i(f^{-i}, \theta_i)$ and $\ell_i(f, \theta_i)$ are equally good for type θ_i. As $s \in \mathcal{B}(\Gamma)$, type θ_i weakly prefers the lottery obtained under message $s_i(\theta_i)$ to the lottery $\ell_i(f, \theta_i)$. As f is interim incentive efficient, f cannot be welfare dominated by $g(s)$, and so $U_i^{g(s)}(\theta_i) = U_i^f(\theta_i)$.

As the choice of type θ_i is arbitrary, we have that $U_i^{g(s)}(\theta_i) = U_i^f(\theta_i)$ for all $i \in I$ and all $\theta_i \in \Theta_i$. As F satisfies closure, it follows from $g(s) \in X^{IC}$ that $g(s) \in F$, as we sought. ∎

References

[1] **d'Aspremont, C., Gérard-Varet, L.-A.** (1979): "Incentives and incomplete information". *Journal of Public Economics* **11**: 25-45.

[2] **d'Aspremont, C., Gérard-Varet, L.-A.** (1982): "Bayesian incentive compatible beliefs". *Journal of Mathematical Economics* **10**(1): 83-103.

[3] **Dasgupta, P., Hammond, P., Maskin, E.** (1979): "The implementation of social choice rules: Some general results on incentive compatibility". *Review of Economic Studies* **46**(2): 185-216.

[4] **Dutta, B., Sen, A.** (1994): "2-person Bayesian implementation". *Economic Design* **1**: 41-54.

[5] **Gresik, T.** (1996): "Incentive efficient equilibria of two-party sealed-bid bargaining games". *Journal of Economic Theory* **68**: 26–48.

[6] **Holmström, B., Myerson, R.** (1983): "Efficient and durable decision rules with incomplete information". *Econometrica* **51**(6): 1799-1819.

[7] **Jackson, M. O.** (1991): "Bayesian implementation". *Econometrica* **59**: 461-477.

[8] **Laussel, D., Palfrey, T.** (2003): "Efficient equilibria in the voluntary contributions mechanism with private information". *Journal of Public Economic Theory* **5**: 449–478.

[9] **Ledyard, J., Palfrey, T.** (1994): "Voting and lottery drafts as efficient public goods mechanisms". *Review of Economic Studies* **61**: 327–355.

[10] **Ledyard, J., Palfrey, T.** (1999a): "Interim efficiency in a public goods problem", in: C. d'Aspremont (Ed.), Social Organization and Mechanism Design, Proceedings of the 1996 Francqui Prize Colloquium, De Boeck University Press, Brussels, Belgium, 1999, pp. 239–261.

[11] **Ledyard, J., Palfrey, T.** (1999b): "A characterization of interim efficiency with public goods". Econometrica **67**: 435–848.

[12] **Ledyard, J., Palfrey, T.** (2002): "The approximation of efficient public good mechanisms by simple voting schemes". *Journal of Public Economic* **83**: 153–172.

[13] **Ledyard, J., Palfrey, T.** (2007): "A general characterization of interim efficient mechanisms for independent linear environments". *Journal of Economic Theory* **133**: 441-466.

[14] **Maskin, E.** (1999): "Nash equilibrium and welfare optimality". *Review of Economic Studies* **66**: 23-38.

[15] **Mookherjee, D., Reichelstein, S.** (1990): "Implementation via augmented revelation mechanisms". *Review of Economic Studies* **57**: 453-476.

[16] **Myerson, R.** (1981): "Optimal Auction Design". *Mathematics of Operations research* 6, 58-73.

[17] **Ollár M, Penta A** (2017): "Full implementation and belief restrictions". *American Economic Review* **107**(8): 2243-2277.

[18] **Palfrey T, Srivastava S** (1986): "Private information in large economies". *Journal of Economic Theory* **39**: 34-58.

[19] **Palfrey, T., Srivastava, S.** (1989a): "Mechanism design with incomplete information: A solution to the implementation problem". *Journal of Political Economy* **97**: 668-691.

[20] **Palfrey, T., Srivastava, S.** (1989b): "Implementation with incomplete information in exchange economies". *Econometrica* **57**: 115-134.

[21] **Palfrey, T., Srivastava, S.** (1991): "Efficient trading mechanisms with preplay communication". *Journal of Economic Theory* **55**: 17-40.

[22] **Postlewaite, A., Schmeidler, D.** (1986): "Implementation in differential information economies". *Journal of Economic Theory* **39**: 14-33.

[23] **Skreta, V.** (2007): "Optimal Auctions with General Distributions. New York University", Stern School of Business, Working Paper.

[24] **Wilson, R.** (1985): "Incentive efficiency of double auctions". *Econometrica* **53**: 1101–1116.