# CCES Discussion Paper Series
## Center for Research on Contemporary Economic Systems

**Graduate School of Economics**
**Hitotsubashi University**

CCES Discussion Paper Series, No.43
July 2011


# Partially-honest Nash implementation: Characterization results


Michele Lombardi
(Maastricht University)
Naoki Yoshihara
(Hitotsubashi University)



Naka 2-1, Kunitachi, Tokyo 186-8601, Japan
Phone: +81-42-580-9076  Fax: +81-42-580-9102
URL: http://www.econ.hit-u.ac.jp/~cces/index.htm
E-mail: cces@econ.hit-u.ac.jp

# Partially-honest Nash implementation: Characterization results*

Michele Lombardi[†] Naoki Yoshihara[‡]

July 20, 2011

**Abstract**

This paper studies implementation problems in the wake of a recent trend of implementation of non-consequentialist nature, which draws on the evidence taken from experimental and behavioral economics. Specifically, following the seminal works by Matsushima (2008) and Dutta and Sen (2009), the paper considers *implementation problems with partially-honest agents*, which presume that there is at least one individual in society who concerns herself with not only outcomes but also honest behavior at least in a limited manner. Given this setting, the paper provides a general characterization of Nash implementation with partially-honest individuals. It also provides the necessary and sufficient condition for Nash implementation with partially-honest individuals by mechanisms with some types of strategy-space reductions. As a consequence, it shows that in contrast to the case of the standard framework, the equivalence between Nash implementation and Nash implementation with strategy space reduction no longer holds.

*JEL classification*: C72; D71.
*Key-words*: Nash implementation, canonical-mechanisms, $s$-mechanisms, self-relevant mechanisms, partial-honesty, permissive results.

*First version: Febraury 13, 2011. We are grateful to Antonio Cabrales, Giulio Codognato, Ahmed Doghmi, Navin Kartik, Rudolf Muller, Hans Peters, Tatsuyoshi Saijo, Olivier Tercieux, Dries Vermeulen, and audiences at the CEPET workshop in Udine, the Fifth Workshop in Decisions, Games & Logic at Maastricht University, the EEA-ESEM joint meeting in Oslo, the SAET conference in Ancao, the 22nd International Conference on Game Theory at Stony Brook University, for useful comments and suggestions.

†Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, Netherlands, phone: 0031 43 3883 761, fax: 0031 43 3882 000, e-mail: m.lombardi@maastrichtuniversity.nl.

‡Institute of Economic Research, Hitotsubashi University, 2-4 Naka, Kunitachi, Tokyo, 186-8603 Japan, phone: 0081 42 580 8354, fax: 0081 42 580 8333, e-mail: yosihara@ier.hit-u.ac.jp.

# 1 Introduction

The theory of (Nash) implementation aims to reach *goals* in situations in which the planner does not have all the relevant necessary information, but needs to elicit it from the agents.[1] To this end, she designs a mechanism or game form in which agents will act strategically in accordance with the solution concept of Nash equilibrium. When the (Nash) equilibrium outcomes of the mechanism coincide with the goals set by the planner, these goals are implementable. A seminal paper on implementation is Maskin (1999), who proves that a social choice correspondence ($SCC$) - which summarizes the planner's goals - is (Maskin) *monotonic* if it is implementable; when there are at least three agents, an $SCC$ is implementable if it is monotonic and satisfies an auxiliary condition called *no-veto power*; this is *Maskin's Theorem*.[2] Moore and Repullo (1990), Dutta and Sen (1991), Danilov (1992), Sjöström (1991), and Yamato (1992) refined Maskin's characterization result by providing necessary and sufficient conditions for an $SCC$ to be implementable.[3]

A fundamental tenet of implementation theory is the consequentialism axiom. Its core idea is that the ranking of outcomes of agents should be independent of the process that generates these outcomes. An immediate implication of this axiom for implementation theory is that agents should be indifferent between a lie and a truthful statement if they result in the same material payoffs.[4] This axiom is, however, inconsistent with the mounting evidence from psychology and economics as well as from causal observations and introspection, that agents may display concern for *procedures*; that is, they may care about *how* outcomes are generated and, therefore, their ranking of outcomes may be structurally dependent on the outcome-generating process (Camerer, 2003; Sen, 1997). Remarkably, a considerable amount of experimental data suggests that agents may display preferences for truth-telling; that is, an agent lies *only* when she prefers the outcome obtained from false-telling over the outcome obtained from truth-telling (Gneezy, 2005; Hurkens and Kartik, 2009). Unexpectedly, these kinds of preferences even emerge in experiments designed to test the feasibility of classical mechanisms for implementation (Cabrales *et al.*, 2003).[5] The paper aims to narrow the gap between these two strands. It follows the non-consequentialist approach by accommodating concerns for truthful revelation of agents; but like mainstream theory, it keeps the idea that even these agents respond primarily to material incentives.[6] The paper refers to agents

[1] Henceforth, by implementation we mean Nash implementation.

[2] The first version appeared in 1977.

[3] For excellent introductions to the theory of implementation, see, for instance, Jackson (2001) and Maskin and Sjöström (2002).

[4] The pioneer work in opening the theory of mechanism design to non-consequentialist considerations is that of Glazer and Rubinstein (1998), where individuals involved in a mechanism care explicitly about the process by which their recommendations affect the social decision, as they desire to see their recommendations coincide with the social choice.

[5] Following the experiment result of Cabrales et. al (2003), the overall rate of truth-telling is 57% in a treatment with no fine. Note that in that paper, a message by agent $i$ is defined to be truthful if she has reported the true state of the world, i.e., for instance, the true preference profiles. Such a definition does not take into account the outcome announcement as a part of agent $i$'s truth-telling strategy.

[6] In its turn, the impressive body of evidence accumulated by psychologists over the past two decades has caused scholars to study the implications of weakening other fundamental assumptions in a variety of ways, and has already turned in a number of alternatives back to the standard implementation model (for instance, Eliaz, 2002; Renou and Schlag, 2009; Bergemann et al., 2010; Cabrales and Serrano, 2010). Noteworthy, the first paper on 'behavioral implementation theory' dates back to 1986, in which Hurwicz solves the implementation problem without positing the completeness and the transitivity of agents' preferences (Hurwicz, 1986).

having preferences for truth-telling as being *partially-honest* or *dishonest averse*.

Its general thrust goes as follows. Assume, as an example, that the message conveyed by each agent to the planner involves the announcement of a preference profile (i.e., agents' preferences over outcomes). A message is truthful if it involves the announcement of the true preference profile. A partially-honest agent is an agent who strictly prefers to announce a truthful message rather than a lie when the former (given a message announced by other agents) produces an outcome which is at least as good as the one that would be achieved if the agent lied (keeping constant the other agents' messages). Suppose that agent $h$ is a partially-honest agent, who believes that the other agents will send the message $m_{-h}$, and let $m_h$ be the truthful message of agent $h$ and $m'_h$ be not truthful. Moreover, let both the message profile $(m_h, m_{-h})$ and the message profile $(m'_h, m_{-h})$ result in the same outcome $x$. Then, unlike an agent who is concerned solely with outcomes, the partially-honest agent $h$ strictly prefers $(m_h, m_{-h})$ to $(m'_h, m_{-h})$. Put differently, the agent at issue has preferences over message profiles in which she cares about two dimensions in lexicographic order: primarily to her outcome, secondarily to her truth-telling behavior.

Seminal works on the role of honesty in implementation theory are Matsushima (2008) and Dutta and Sen (2009), which show that the assumption that the planner is aware of the existence of partially-honest agents but ignores their identities drastically improves the scope of implementation. Yet, the significant impact of the presence of partially-honest agents upon implementation theory has not been fully appreciated - as described below. In line with these works, this paper also investigates implementation problems with partially-honest agents, where an $SCC$ is *partially-honest implementable* if there is a mechanism whose equilibrium outcomes are determined with each profile of preferences over message profiles as well as potential sets of partially honest agents, and coincide with the optimal outcomes set by this $SCC$.

Given this setting, the paper provides, in section 3.1, a minimal set of necessary conditions for partially-honest implementation, though the above seminal works solely study sufficient conditions. Due to this result in the paper, it is possible to examine which of the $SCC$s cannot be partially-honestly implemented. For instance, as shown in section 4, the (strong) $Pareto$ $SCC$ defined in abstract social choice environments is not partially-honestly implementable. Furthermore, under mild and reasonable domain restrictions of preferences and mechanisms, the paper shows that a slight strengthening of these conditions is necessary and sufficient for partially-honest implementation in more than two person societies. The set of conditions is much weaker than the necessary and sufficient condition given by Moore and Repullo (1990) for the standard implementation, and in particular it contains no variant of the Maskin monotonicity-like condition. For instance, in rationing problems when agents have single-plateaued preferences, this characterization shows that the $Pareto$ $SCC$ is partially-honest implementable, though this $SCC$ violates the Moore and Repullo (1990) condition, and also satisfies neither monotonicity nor no-veto power.

Note that the aforementioned theorem of this paper applies a canonical mechanism to show the sufficiency part. This type of mechanism requests agents to announce a feasible social outcome, an agent index, and moreover a profile of agents' preferences on outcomes, which is not an attractive feature, given that an important role of the mechanism is to economize on communication. Facing this issue, the paper pays attention to informational decentralization of mechanisms by considering mechanisms with strategy space reductions. While sub-section 3.2 assumes $s$-mechanisms (Saijo, 1988) in which the message conveyed by each agent to the planner involves the announcement of only her own and her neighbor's preferences - in addition to an outcome and an agent index, sub-section 3.3 endorses the idea of self-relevant mechanisms (Tatamitani, 2001) according to which each agent announces - *inter alia* - only her own preference. Then, the paper identifies a minimal set of necessary conditions for partially-honest implementation by $s$-mechanisms (*resp.*, self-relevant mechanisms); moreover, it shows that a slight strengthening of these conditions fully identifies the class of partially-honest implementable $SCC$s by $s$-mechanisms (*resp.*, self-relevant mechanisms). Notably, these conditions respectively contain weaker variants of (Maskin) monotonicity-type conditions, each of which respectively restricts the class of partially-honest implementable $SCC$s by $s$-mechanisms and by self-relevant mechanisms. These findings have at least two immediate consequences. First, there is a trade-off between what the planner can achieve when there are partially-honest agents in the society and the strengthening of informational decentralization in mechanisms. Second, this conflict breaks down the equivalence between implementation and implementation by $s$-mechanism which holds in the standard framework (Lombardi and Yoshihara, 2010).

Finally, the paper turns to study partially-honest implementation problems in two-agent societies. This issue has recently been analyzed by Dutta and Sen (2009) on the assumption that agents' preferences are linear orders. Their contribution is that, even in the more problematic case of two agents, the stringent condition of monotonicity is no longer required. The paper extends their analysis to the domain of weak orders in view of its potential applications to bargaining and negotiating. The paper identifies the class of partially-honest implementable $SCC$s, not only in the case that the planner knows that exactly one agent is partially-honest, but also in the more subtle case that she only knows that there exist partially-honest agents.

The paper is organized as follows. Section 2 describes the formal environment. Section 3 reports the analysis for the many-person case, whereas Section 4 discusses briefly its implications. Section 5 reports the analysis for the two-agent case and its implications. Section 6 concludes briefly.

# 2 The implementation problem

The set of outcomes is denoted by $X$ and the set of agents is $N = \{1, ..., n\}$. Unless otherwise specified, we assume that the cardinality of $X$ is $\#X \geq 2$, while the cardinality of $N$ is $n \geq 3$. Let $\mathcal{R}(X)$ be the set of all possible weak orders on $X$.[7] Let $\mathcal{R}_\ell \subseteq \mathcal{R}(X)$ be the (non-empty) set of all admissible weak orders for agent $\ell \in N$.[8] Let $\mathcal{R}^n \subseteq \mathcal{R}_1 \times ... \times \mathcal{R}_n$ be the set of all admissible profiles of weak orders (or states). A generic element of $\mathcal{R}^n$ is denoted by $R$, where its $\ell$th component is $R_\ell \in \mathcal{R}_\ell$, $\ell \in N$.[9] The symmetric and asymmetric factors of any $R_\ell \in \mathcal{R}_\ell$ are, in turn, denoted $P_\ell$ and $I_\ell$, respectively.[10] For any $R \in \mathcal{R}^n$ and any $\ell \in N$, let $R_{-\ell}$ be the list of elements of $R$ for all agents except $\ell$, i.e., $R_{-\ell} \equiv (R_1, ..., R_{\ell-1}, R_{\ell+1}, ..., R_n)$. Given a list $R_{-\ell}$ and $R_\ell \in \mathcal{R}_\ell$, we denote by $(R_{-\ell}, R_\ell)$ the preference profile consisting of these $R_\ell$ and $R_{-\ell}$. For any preference profile $R \in \mathcal{R}^n$ and any $\varnothing \neq S \subseteq N$, let $R_{-S}$ be the

[7] A weak order is a complete and transitive binary relation. A relation $R$ on $X$ is *complete* if, for all $x, x' \in X$, $(x, x') \in R$ or $(x', x) \in R$; *transitive* if, for all $x, x', x'' \in X$, if $(x, x') \in R$ and $(x', x'') \in R$ , then $(x, x'') \in R$.

[8] The weak set inclusion is denoted by $\subseteq$, while the strict set inclusion is denoted by $\subsetneq$.

[9] $(x, y) \in R_\ell$ stands for "$x$ is at least as good as $y$".

[10] $(x, y) \in P_\ell$ if and only if $(x, y) \in R_\ell$ and $(y, x) \notin R_\ell$ and $P_\ell$ stands for "strictly better than". On the other hand, $(x, y) \in I_\ell$ if and only if $(x, y) \in R_\ell$ and $(y, x) \in R_\ell$ and $I_\ell$ stands for "indifferent to".

list of elements of $R$ for all agents in $N \backslash S$. Given a list $R_{-S}$ and a list $R_S \in \times_{\ell \in S} \mathcal{R}_\ell$, we denote by $(R_{-S}, R_S)$ the preference profile consisting of these $R_S$ and $R_{-S}$. Let $\mathcal{P}^n \subseteq \mathcal{R}^n$ be the set of all admissible profiles of linear orders.[11] Let $L(R_\ell, x)$ denote agent $i$'s lower contour set at $(R_\ell, x) \in \mathcal{R}_\ell \times X$, that is, $L(R_\ell, x) \equiv \{y \in X | (x, y) \in R_\ell\}$. For any $R_\ell \in \mathcal{R}_\ell$ and $Y \subseteq X$, let $\max_{R_\ell} Y$ be the set of optimal outcomes in $Y$ according to $R_\ell$, that is, $\max_{R_\ell} Y \equiv \{x \in Y | (x, y) \in R_\ell \text{ for all } y \in Y\}$. For any $(R_\ell, x) \in \mathcal{R}_\ell \times X$, $\partial L(R_\ell, x) = \{x\}$ means $\{x\} = \max_{R_\ell} L(R_\ell, x)$.

A *social choice correspondence* (*SCC*) $F$ on $\mathcal{R}^n$ is a correspondence $F : \mathcal{R}^n \twoheadrightarrow X$ with $\varnothing \neq F(R) \subseteq X$ for all $R \in \mathcal{R}^n$. Denote the class of such correspondences by $\mathcal{F}$. An *SCC* $F$ on $\mathcal{R}^n$ is (Maskin) *monotonic* if, for all $R, R' \in \mathcal{R}^n$, with $x \in F(R)$, $x \in F(R')$ holds whenever $L(R_\ell, x) \subseteq L(R'_\ell, x)$ for all $\ell \in N$. An *SCC* $F$ on $\mathcal{R}^n$ satisfies i) *no-veto power* if, for all $R \in \mathcal{R}^n$, $x \in F(R)$ holds whenever $x \in \max_{R_\ell} X$ for at least $n-1$ agents; ii) *unanimity* if, for all $R \in \mathcal{R}^n$, $x \in F(R)$ holds whenever $x \in \max_{R_\ell} X$ for all $\ell \in N$. Given an *SCC* $F$, an outcome $x$ is $F$-optimal at a preference profile $R \in \mathcal{R}^n$ if $x \in F(R)$.

A *mechanism* or *game form* is a pair $\gamma \equiv (M, g)$, where $M \equiv M_1 \times ... \times M_n$, with each $M_i$ being a (non-empty) set, and $g : M \to X$. It consists of a message space $M$, where $M_\ell$ is the message space for agent $\ell \in N$, and an outcome function $g$. Denote the admissible class of mechanisms by $\Gamma$. Let $m_\ell \in M_\ell$ denote a generic message (or strategy) for agent $\ell$. A message profile is denoted by $m \equiv (m_1, ..., m_n) \in M$. For any $m \in M$ and $\ell \in N$, let $m_{-\ell} \equiv (m_1, ..., m_{\ell-1}, m_{\ell+1}, ..., m_n)$. Let $M_{-\ell} \equiv \times_{i \in N \backslash \{\ell\}} M_i$. Given an $m_{-\ell} \in M_{-\ell}$ and an $m_\ell \in M_\ell$, denote by $(m_\ell, m_{-\ell})$ the message profile consisting of these $m_\ell$ and $m_{-\ell}$. For any $m \in M$ and $\varnothing \neq S \subseteq N$, let $m_{-S} \equiv (m_\ell)_{\ell \in N \backslash S}$. Let $M_{-S} \equiv \times_{\ell \in N \backslash S} M_\ell$. Given $m_{-S} \in M_{-S}$ and $m_S \in M_S$, denote by $(m_S, m_{-S})$ the message profile consisting of these $m_S$ and $m_{-S}$.

A mechanism $\gamma$ induces a class of (*non-cooperative*) *games* $\{(\gamma, R) | R \in \mathcal{R}^n\}$. Given a game $(\gamma, R)$, we say that $m^* \in M$ is a (pure strategy) *Nash equilibrium* at $R$ if and only if, for all $\ell \in N$, $\left(m^*, \left(m_\ell, m^*_{-\ell}\right)\right) \in R_\ell$ for all $m_\ell \in M_\ell$. Given a game $(\gamma, R)$, let $NE(\gamma, R)$ denote the set of Nash equilibrium message profiles of $(\gamma, R)$, whereas $NA(\gamma, R)$ represents the corresponding set of Nash equilibrium outcomes.

A mechanism $\gamma$ *implements* $F$ in *Nash equilibria*, or simply *implements* $F$, if and only if $F(R) = NA(\gamma, R)$ for all $R \in \mathcal{R}^n$. If such a mechanism exists, then $F$ is (*Nash*)-*implementable*.

Given a mechanism $\gamma$, for each agent $\ell \in N$ a *truth-telling correspondence* $T^\gamma_\ell$ on $\mathcal{R}^n \times \mathcal{F}$ is a correspondence $T^\gamma_\ell : \mathcal{R}^n \times \mathcal{F} \twoheadrightarrow M_\ell$ with $\varnothing \neq T^\gamma_\ell(R, F) \subseteq M_\ell$ for each $(R, F) \in \mathcal{R}^n \times \mathcal{F}$. An interpretation of the set $T^\gamma_\ell(R, F)$ is that, given the mechanism $\gamma$ and the pair $(R, F)$, agent $\ell$ behaves truthfully at the message profile $m \in M$ if and only if $m_\ell \in T^\gamma_\ell(R, F)$. In other words, $T^\gamma_\ell(R, F)$ is the set of *truthful messages* of agent $\ell$ under the mechanism $\gamma$, when the current social state is $R \in \mathcal{R}^n$ and the social goal is given by $F$. Note that the type of elements of $M_\ell$ constituting $T^\gamma_\ell(R, F)$ depends on the type of mechanism $\gamma$ that one may consider. For example, if the message conveyed by each agent to the planner involves the announcement of a preference profile, a feasible outcome and an agent index, and sending the truthful preference profile constitutes the relevant truthful message for each $(R, F) \in \mathcal{R}^n \times \mathcal{F}$, then $M_\ell$ may be defined by $M_\ell \equiv M^1_\ell \times M^2_\ell$, where there is a bijection $\sigma_\ell : \mathcal{R}^n \to M^1_\ell$ such that $T^\gamma_\ell(R, F) = \{\sigma_\ell(R)\} \times M^2_\ell$ for each $(R, F) \in \mathcal{R}^n \times \mathcal{F}$.

For any $\ell \in N$ and $R \in \mathcal{R}^n$, let $\succcurlyeq^R_\ell$ be agent $\ell$'s weak order over $M$ under the state $R$. The asymmetric factor of $\succcurlyeq^R_\ell$ is denoted $\succ^R_\ell$, while the symmetric part is denoted $\sim^R_\ell$. For

[11] A linear order is a complete, transitive, and antisymmetric binary relation. A binary relation $R$ on $X$ is *antisymmetric* if, for all $x, x' \in X$, $x = x'$ if $(x, x') \in R$ and $(x', x) \in R$.

any $R \in \mathcal{R}^n$, let $\succcurlyeq^R$ denote the profile of weak orders over $M$ under the state $R$, that is, $\succcurlyeq^R \equiv \left(\succcurlyeq^R_\ell\right)_{\ell \in N}$.

**Definition 1**. An agent $h \in N$ is a *partially-honest* or *dishonest averse agent* if, for any mechanism $\gamma$, any $R \in \mathcal{R}^n$, and any $m \equiv (m_h, m_{-h}), m' \equiv (m'_h, m_{-h}) \in M$, the following properties hold:
(i) if $m_h \in T^\gamma_h(R, F)$, $m'_h \notin T^\gamma_h(R, F)$, and $(g(m), g(m')) \in R_h$, then $(m, m') \in \succ^R_h$;
(ii) otherwise, $(m, m') \in \succcurlyeq^R_h$ if and only if $(g(m), g(m')) \in R_h$.

An agent $\ell \in N$ who is a partially-honest agent is denoted by $h$. If agent $\ell \in N$ is not a partially-honest agent, i.e., $\ell \neq h$, then for each game $(\gamma, R)$, for all $m, m' \in M$: $(m, m') \in \succcurlyeq^R_\ell$ if and only if $(g(m), g(m')) \in R_\ell$.

Unless otherwise specified, the following informational assumption holds throughout the paper.

**Assumption 1.** *There are* partially-honest agents *in $N$. The planner is well aware of the fact that there are partially-honest agents in $N$ but she does not know their identities.*

Thus, while the planner knows that there are partially-honest agents in society and how these agents behave, the planner knows neither the identity of the partially-honest agents nor their exact number.

Let $\mathcal{H} \subseteq \{H \subseteq N \mid H \neq \varnothing\}$ be the class of subsets of $N$. Note that $\mathcal{H}$ is considered as the potential class of partially-honest agents' groups. That is, if $H \in \mathcal{H}$, this $H$ is a *potential group of partially-honest agents* in $N$; in other words, $H$ is a conceivable set of partially-honest agents. By Assumption 1, the planner knows that $\mathcal{H}$ is non-empty, and perhaps, she may know what subsets of $N$ belong to $\mathcal{H}$, but she never knows which element of $\mathcal{H}$ is the true set of partially-honest agents in the society. Assumption 1 implies that $\#\mathcal{H} \geq 2$.

A mechanism $\gamma$ induces a class of (*non-cooperative*) *games with partially-honest agents* $\left\{\left(\gamma, \succcurlyeq^R\right) | R \in \mathcal{R}^n, H \in \mathcal{H}\right\}$. Given a game $\left(\gamma, \succcurlyeq^R\right)$, we say that $m^* \in M$ is a (pure strategy) *Nash equilibrium with partially-honest agents* at $(R, H)$ if and only if, for all $\ell \in N$, $\left(m^*, \left(m_\ell, m^*_{-\ell}\right)\right) \in \succcurlyeq^R_\ell$ for all $m_\ell \in M_\ell$. Given a game $\left(\gamma, \succcurlyeq^R\right)$, let $NE\left(\gamma, \succcurlyeq^R\right)$ denote the set of Nash equilibrium message profiles of $\left(\gamma, \succcurlyeq^R\right)$, whereas $NA\left(\gamma, \succcurlyeq^R\right)$ represents the corresponding set of Nash equilibrium outcomes.

Since by Assumption 1 the planner knows that there are partially-honest agents in $N$ but not who these agents are, this raises the question of what is an appropriate notion of implementation in such a setting. To enable the planner to partially-honestly implement $SCC$s, the paper amends the standard definition of implementation as follows.

**Definition 2.** An $SCC$ $F \in \mathcal{F}$ is *partially-honest (Nash) implementable* if there exists a mechanism $\gamma = (M, g) \in \Gamma$ such that $F(R) = NA\left(\gamma, \succcurlyeq^R\right)$ for all $R \in \mathcal{R}^n$ and all $H \in \mathcal{H}$.

In the conventional implementation theory, the objective of the planner is to design a mechanism whose equilibrium outcomes coincide with the $F$-optimal outcomes for each admissible state $R$. In contrast, in the presence of partially-honest agents, the planner, to achieve the implementability of the goal $F$, has to design a mechanism in which the equivalence between the set of equilibrium outcomes and the set of $F$-optimal outcomes holds not only for each admissible state $R$, but also for each conceivable set of partially-honest agents, i.e., for each $H \in \mathcal{H}$. Note that the gap between the two definitions becomes closed when no agent in $N$ is partially-honest.

To conclude, let us introduce two mild conditions imposed on the models of this paper. One is a condition on the domain of agents' preferences, while the other is a condition on the domain of mechanisms admissible in the society. The first condition basically requires that the class of available profiles of preferences is sufficiently rich. Examples of preference domains satisfying such a condition would be the set of all profiles of weak orders, linear orders, and single peaked preferences on $X$. Moreover, it is vacuously satisfied in the classical economic environments. Hence, our models are applicable to those environments. The condition can be stated as follows.

**Rich Domain (RD)**: For any $i \in N$, any $R \in \mathcal{R}^n$, and any $x \in X$, if $R'_i \in \mathcal{R}_i(X)$ is such that $L(R'_i, x) = L(R_i, x)$ with $\partial L(R'_i, x) = \{x\}$, then $(R'_i, R_{-i}) \in \mathcal{R}^n$ holds.

Next, our informational assumption is that the planner knows that there exist partially-honest agents but ignores their identities. The partially-honest agent is an agent who prefers to be truthful if a lie is not beneficial to her. Given this structure, the existence of truthful messages is presumed since otherwise, the issue reduces to the standard implementation problem. Moreover, the admissible class of mechanisms should be constituted by those which involve a *simple scheme* to punish such a partially-honest agent if she sends a false message. Within this class, let us consider a type of mechanism in which, if an outcome $x$ is $F$-optimal at the state $R$ and the outcome function $g$ selects $x$ as the resulting outcome of the messages announced by agents, a partially-honest agent can find a truthful message which results in the same outcome $x$ - keeping constant the messages of all other agents. In such a mechanism, any false statement by a partially-honest agent can be punished independently of the detailed information about the true state of the society. The condition on the class of admissible mechanisms $\Gamma$ can be stated as follows.

**Simple Punishment (SP):** For any $F \in \mathcal{F}$, for any $R, R' \in \mathcal{R}^n$, any $x \in F(R)$, any $i \in N$, and any $m \in M$ such that $g(m) = x$, there is $m'_i \in T_i^{\gamma}(R', F)$ such that $g(m'_i, m_{-i}) = g(m)$.

A mechanism $\gamma$ is a *mechanism with simple punishment* if it satisfies **SP**. Denote the class of mechanisms with **SP** by $\Gamma_{SP}$.

Before closing this section, it may be worth noting that the simple punishment property is satisfied by all classical mechanisms in the literature of Nash implementation (see, for instance, Repullo, 1987; Moore and Repullo, 1990; Saijo, 1988; Dutta and Sen, 1991; Tatamitani, 2001).

# 3 Characterization results for the many-person case

This section reports the analysis of partially-honest implementation problems in the many-person case.

Sub-section 3.1 basically imposes no restriction on the types of admissible mechanisms except for $\Gamma = \Gamma_{SP}$. First, this sub-section identifies a minimal set of necessary conditions for partially-honest implementation with no restriction on $\Gamma$. The necessary conditions include only weaker variants of the no-veto power condition. Then, by setting $\Gamma = \Gamma_{SP}$, it is shown that a slight strengthening of this minimal set of necessary conditions fully identifies the class of $SCC$s that are partially-honest implementable - by *canonical mechanisms.*

The section, then, turns to study partially-honest implementation by mechanisms with strategy space reductions. While sub-section 3.2 studies partially-honest implementation by *s-mechanisms*, sub-section 3.3 analyzes the same implementation problem by focussing on *self-relevant mechanisms.* For each of these two types of mechanisms, their respective sub-sections identify a minimal set of necessary conditions that an $SCC$ $F$ must satisfy if it is partially-honest implementable. The identified necessary conditions incorporate a Maskin monotonicity-like condition. Finally, given $\Gamma = \Gamma_{SP}$, it is reported that a slight strengthening of the necessary conditions for $s$-mechanisms (*resp.*, self-relevant mechanisms) fully characterizes partially-honest implementation by $s$-mechanisms (*resp.*, self-relevant mechanisms).

The sets of conditions that are necessary and sufficient for partially-honest implementation are more complex than those obtained by Moore and Repullo (1990), Tatamitani (2001), and Lombardi and Yoshihara (2010), but they are remarkably weaker and do provide additional insights; we refer the reader to Section 4 for more details.

## 3.1 Partially-honest implementation: A general characterization

Since Maskin's Theorem, there have been impressive advances in implementation theory. Specifically, in societies with at least three agents, Moore and Repullo (1990) established that an *SCC* $F$ is implementable if and only if it satisfies Condition $\mu$ defined below.

Condition $\mu$ (for short, $\mu$): There is a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there is a profile of sets $(C_\ell(R,x))_{\ell \in N}$ such that $x \in C_\ell(R,x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; finally, for all $R^* \in \mathcal{R}^n$, the following (i)-(iii) are satisfied:
(i) if $C_\ell(R,x) \subseteq L(R^*_\ell, x)$ for all $\ell \in N$, then $x \in F(R^*)$;
(ii) for all $i \in N$, if $y \in C_i(R,x) \subseteq L(R^*_i, y)$ and $y \in \max_{R^*_\ell} Y$ for all $\ell \in N \setminus \{i\}$, then $y \in F(R^*)$;
(iii) if $y \in \max_{R^*_\ell} Y$ for all $\ell \in N$, then $y \in F(R^*)$.[12]

Condition $\mu$(i) is equivalent to monotonicity, while Conditions $\mu$(ii) and $\mu$(iii) are weaker versions of no-veto power.

Our first task in this sub-section is to find necessary conditions for an *SCC* to be partially-honest implementable. These conditions will not include any monotonicity-type condition, since the injection of a minimal dishonest aversion into implementation theory frees us from the jail of Maskin monotonicity (see Dutta and Sen, 2009). Yet, this task is particularly complicated and subtle when indifference relations are allowed.[13] To explain this aspect, suppose that an *SCC* $F$ is partially-honest implementable by a mechanism $\gamma$. Let $Y$ be the range of $g$:

$$Y \equiv g(M) = \{x \in X | g(m) = x \text{ for some } m \in M\}.$$

Consider a preference profile $R^* \in \mathcal{R}^n$. Suppose that some outcome $y = g(m)$ in $Y$ is an optimal outcome under the state $R^*$ in the set $Y$ for all agents so as to fulfill the premises of Condition $\mu$(iii). In the conventional theory, the message profile $m$ constitutes an equilibrium of the game $(\gamma, R^*)$. However, it may not be the case when there are partially-honest agents. For the sake of simplicity, assume that only agent $h$ is partially-honest. Suppose that the message $m_h$ conveyed to the planner by this agent is not truthful, while a truthful statement, say $m'_h \in T^\gamma_h(R^*, F)$, results in an outcome $x = g(m'_h, m_{-h})$ distinct from $y$ for which she is indifferent to. Suppose that $x$ is not maximal for one of the other agents. In this situation, we can no longer conclude that the outcome $y$ is *SCC*-optimal, as the message profile $m$

[12] We refer to the condition that requires only one of the conditions (i)–(iii) in Condition $\mu$ as Conditions $\mu$(i)–$\mu$(iii) each. Note that Condition $\mu$ implies Conditions $\mu$(i)–$\mu$(iii), but the converse is not true. We use similar conventions below.

[13] When the domain of preferences contains only linear orders, Condition $\mu$ without Condition $\mu$(i) is not only necessary but sufficient.

supporting $y$ is not an equilibrium of the game $\left(\gamma, \succcurlyeq^{R^*}\right)$ - since agent $h$ strictly prefers $(m'_h, m_{-h})$ to $m$. This indicates that even when $y$ is maximal in $Y$ under $R^*$, not all strategies in $g^{-1}(y)$ can constitute an equilibrium of $g$ at $R^*$ when there are partially-honest agents. Among these strategies, only those in which all partially-honest agents are making truthful reports may support $y$ as an $F$-optimal outcome at $R^*$. This can be achieved by requiring that for *all potential partially-honest* agents (since the identities of partially-honest agents are unknown), the outcome $y$ must be the unique optimal outcome under $R^*$ in the set $Y$. With this additional requirement, agent $h$ can profitably deviate from $m_h \notin T_h^\gamma(R^*, F)$ to an $m''_h \in T_h^\gamma(R^*, F)$, but her deviation will not prevent us from concluding that $y$ is $F$-optimal at $R^*$, since the strategy profile $(m''_h, m_{-h})$, when executed by $g$, results in the outcome $y$.

The complications associated with necessary conditions are not limited to Condition $\mu$(iii). The difficulties come mainly from two causes. First, the presence of partially-honest agents breaks down the equivalent relationship between agents' preferences over outcomes and their preferences over message profiles, which is implicitly assumed in the conventional theory. Second, conditions on $F$ are to be formulated only in terms of preferences over outcomes. Taking these difficulties into account, we obtain the following condition, Condition $\mu^*$, which basically contains only weaker versions of Conditions $\mu$(ii) and $\mu$(iii).

CONDITION $\mu^*$ (for short, $\mu^*$): There is a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there is a profile of sets $(C_\ell(R,x))_{\ell \in N}$ such that $x \in C_\ell(R,x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; finally, for all $H \in \mathcal{H}$ and all $R^* \in \mathcal{R}^n$, the following (i)-(iii) are satisfied:
(i) if $C_\ell(R,x) \subseteq L(R^*_\ell, x)$ for all $\ell \in N$ and $x \notin F(R^*)$, then there exists $h \in H$ such that $(x, x') \in I^*_h$ for some $x' \in C_h(R,x)$;
(ii) for all $i \in N$, if $y \in C_i(R,x) \subseteq L(R^*_i, y)$, $y \in \max_{R^*_\ell} Y$ for all $\ell \in N \setminus \{i\}$, and $y \notin F(R^*)$, then:
(a) if $H = \{i\}$, then $(y, y') \in I^*_i$ for some $y' \in C_i(R,x) \setminus \{y\}$;
(b) if $i \in H$ and $\#H > 1$, then $R^* \neq R$ or $(y, y') \in I^*_i$ for some $y' \in C_i(R,x) \setminus \{y\}$;
(iii) if $y \in \max_{R^*_\ell} Y$ for all $\ell \in N$ and $y \notin F(R^*)$, then there is an $h \in H$ such that $(y, y') \in I^*_h$ for some $y' \in Y \setminus \{y\}$.

Notice that Condition $\mu^*$(i) imposes a requirement which is met by all *SCC*s.

The following theorem shows that Condition $\mu^*$ is a minimal set of necessary conditions for the partially-honest implementation.

**Theorem 1.** *Let Assumption 1 hold. If an SCC $F \in \mathcal{F}$ is* partially-honest implementable, *then it satisfies* Condition $\mu^*$.

**Proof.** Let Assumption 1 hold. Let $\gamma \equiv (M, g)$ be a mechanism which partially-honest implements $F \in \mathcal{F}$. Let $Y \equiv g(M)$. Take any $H' \in \mathcal{H}$, $R \in \mathcal{R}^n$, and $x \in F(R)$. Then, there is a strategy $m^{H'} \in NE\left(\gamma, \succcurlyeq^R\right)$ such that $g\left(m^{H'}\right) = x$. Then, $\{x\} \subseteq g\left(M_\ell, m^{H'}_{-\ell}\right) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$. Let $C^{H'}_\ell(R,x) \equiv g\left(M_\ell, m^{H'}_{-\ell}\right)$ for all $\ell \in N$. Define $C_\ell(R,x) \equiv \cup_{H' \in \mathcal{H}} C^{H'}_\ell(R,x)$ for all $\ell \in N$, all $R \in \mathcal{R}^n$, and all $x \in F(R)$. Then, $x \in C_\ell(R,x) \subseteq L(R_\ell, x) \cap Y$ holds for all $\ell \in N$, all $R \in \mathcal{R}^n$, and all $x \in F(R)$. Take any $(R^*, H) \in \mathcal{R}^n \times \mathcal{H}$.

As it is easy to see that $F$ satisfies Condition $\mu^*$(i), we omit the proof here. Next, we show that $F$ meets conditions $\mu^*$(ii)-$\mu^*$(iii).

Pick any $i \in N$ and suppose that $y \in C_i(R,x) \subseteq L(R^*_i, y)$, $y \in \max_{R^*_\ell} Y$ for all $\ell \in N \setminus \{i\}$, and $y \notin F(R^*)$. Then, as $F$ is partially-honestly implemented by $\gamma$, it follows that $y \notin NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for all $H' \in \mathcal{H}$. Since $C_i(R,x) = \cup_{H' \in \mathcal{H}} g\left(M_i, m^{H'}_{-i}\right)$, there exists an $m^{H'} \in NE\left(\gamma, \succcurlyeq^R\right)$ for some $H' \in \mathcal{H}$, such that $g\left(m^{H'}\right) = x$ and $g\left(m'_i, m^{H'}_{-i}\right) = y$ for some

$m_i' \in M_i$. Let $\hat{m} \equiv \left(m_i', m_{-i}^{H'}\right)$. Note that $g(\hat{m}) = y \notin NA\left(\gamma, \succcurlyeq^{R^*}\right)$ holds for any $H' \in \mathcal{H}$. However, since $g(M_i, \hat{m}_{-i}) \subseteq C_i(R,x) \subseteq L(R^*, y)$ and $y \in \max_{R_\ell^*} g(M)$ for all $\ell \in N \setminus \{i\}$ by the premise of $\mu^*$(ii), $\hat{m} \in NE(\gamma, R^*)$ holds. In contrast, the fact that $y \notin NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for any $H' \in \mathcal{H}$ implies that $\hat{m} \notin NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for any $H' \in \mathcal{H}$. Thus, for each $H' \in \mathcal{H}$, there should be an $h \in H'$ such that $\hat{m}_h \notin T_h^\gamma(R^*, F)$ and $(g(m_h^*, \hat{m}_{-h}), g(\hat{m})) \in I_h^*$ for some $m_h^* \in T_h^\gamma(R^*, F)$, since $\hat{m} \in NE(\gamma, R^*) \setminus NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for all $H' \in \mathcal{H}$.

Let $H = \{i\}$, and assume, to the contrary, that $\{y\} = \max_{R_i^*} C_i(R,x)$. Then, $g(m_i^*, \hat{m}_{-i}) = g(\hat{m})$, where $m_i^* \in T_i^\gamma(R^*, F)$ and $\hat{m}_i \notin T_i^\gamma(R^*, F)$ are such that $(g(m_i^*, \hat{m}_{-i}), g(\hat{m})) \in I_i^*$ for the unique partially-honest agent $\{i\} = H$. Since there cannot be any profitable deviation from $(m_i^*, \hat{m}_{-i})$, we have that $y \in NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for some $H = \{i\}$, a contradiction. Thus, $F$ satisfies $\mu^*$(ii.a).

Let $\#H > 1$ and $i \in H$. Suppose $R^* = R$. Then, $(x, y) \in I_i^*$ with $x \neq y$. Note that if $R^* = R$ and $x = y$, then $x \notin NA\left(\gamma, \succcurlyeq^{R}\right)$ for all $H \in \mathcal{H}$, which is a contradiction. Thus, if $R^* = R$, then $(y, y') \in I_i^*$ for some $y' \in C_i(R,x) \setminus \{y\}$, since $x \in C_i(R,x)$. Therefore, $F$ satisfies $\mu^*$(ii.b).

Finally, we show that $F$ satisfies condition $\mu^*$(iii). Let $y \in \max_{R_\ell^*} Y = \max_{R_\ell^*} g(M)$ for all $\ell \in N$, and $y \notin F(R^*)$. As $F$ is partially-honestly implemented by $\gamma$, it follows that $y \notin NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for all $H' \in \mathcal{H}$. Then, $g(\hat{m}) = y$ for some $\hat{m} \in M$. Assume, to the contrary, $\{y\} = \max_{R_\ell^*} g(M)$ for all $h \in H$. As $\hat{m} \notin NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for all $H' \in \mathcal{H}$ and $y \in \max_{R_\ell^*} g(M)$ for all $\ell \in N$, the only agents that could profitably deviate from $\hat{m}$ are the agents in the set $H$. Let $\bar{H} \subseteq H$ be the set of all partially-honest agents $h$ such that $\hat{m}_h \notin T_h^\gamma(R^*, F)$. Consider the profile of profitable deviations $m_{\bar{H}} \equiv (\bar{m}_h)_{h \in \bar{H}}$ such that $\bar{m}_h \in T_h^\gamma(R^*, F)$ for all $h \in \bar{H}$. As $\{y\} = \max_{R_\ell^*} g(M)$ for all $\ell \in H$, we have that $g(\bar{m}_{\bar{H}}, \hat{m}_{-\bar{H}}) = y$. Since there cannot be any profitable deviation from $(\bar{m}_{\bar{H}}, \hat{m}_{-\bar{H}})$, we have that $y \in NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for the set $H$, which is a contradiction. Therefore, $F$ satisfies $\mu^*$(iii). ■

Note that when $\{\{1\}, \{2\}, \ldots, \{n\}\} \subseteq \mathcal{H}$, then $\mu^*$(iii) implies the following:

(iii) if $y \in \max_{R_\ell^*} Y$ for all $\ell \in N$ and $y \notin F(R^*)$, then for any $h \in N$, there exists $y' \in Y \setminus \{y\}$ such that $(y, y') \in I_h^*$.

This is because $y \notin F(R^*)$ implies $y \notin NA\left(\gamma, \succcurlyeq^{R^*}\right)$ in the game at $(R^*, H)$ for all $H \in \mathcal{H}$.

Condition $\mu^*$ alone is not a sufficient condition for partially-honest implementation, but it is sufficient together with some auxiliary conditions if the domain of preferences is sufficiently rich. Such a slightly strengthened condition can be stated as follows.

Condition $\mu^{**}$ (for short, $\mu^{**}$): There is a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there is a profile of sets $(C_\ell(R,x))_{\ell \in N}$ such that $x \in C_\ell(R,x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; Condition $\mu^*$ and Condition $\mu$(iii) hold;[14] finally, for all $H \in \mathcal{H}$, and for all $R^* \in \mathcal{R}^n$, the following conditions are satisfied for all $i \in N$:

(ii.c) if $y \in C_i(R,x) \subseteq L(R_i^*, y)$, $y \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, and $y \notin F(R^*)$, then $[i \notin H \Rightarrow R \neq R^*]$;

(iv) if $L(R_i^*, x) = L(R_i, x)$, $x \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, $R_{-i}^* = R_{-i}$, and $x \notin F(R^*)$, then $H \neq \{i\}$.

Assuming that only mechanisms with simple punishment are admissible, Condition $\mu^{**}$ is necessary and sufficient for partially-honest implementation. Before stating our second main result, it may be instructive to briefly discuss the devised implementing mechanism.

[14] Henceforth, Condition $\mu$(iii) is referred to as Condition $\mu^{**}$(iii). Moreover, we refer to the statement that requires only one of the statements (i) and (ii) in Condition $\mu^*$ as Conditions $\mu^{**}$(i) and $\mu^{**}$(ii).

Let $\gamma = (g, M)$ be a mechanism where for each agent $i \in N$ the message space is $M_i \equiv \mathcal{R}^n \times Y \times N$, with $Y \subseteq X$.[15] Thus, each agent $i$ announces a preference profile, $R^i$, an outcome, $x^i$, and an agent index, $k^i$. Since a central ingredient of our implementation model is that there is a minimal degree of honesty among agents involved in the mechanism $\gamma$, we shall define accordingly what constitutes a honest message for $\gamma$. By endorsing the idea of Dutta and Sen (2009), a message by agent $i$ is truthful for the mechanism $\gamma$ if it discloses to the planner the true preferences of all agents involved in it. Formally, for each $i \in N$, the set of truth-telling messages is

$$T_i^{\gamma}(R, F) \equiv R \times Y \times N \tag{1}$$

for any state, $R \in \mathcal{R}^n$, and any societal goal, $F \in \mathcal{F}$. Finally, let us define the outcome function $g$ as follows. For any message profile $m \in M$,

*Rule 1*: If $(R^\ell, x^\ell) = (\bar{R}, x)$ for all $\ell \in N$ and $x \in F(\bar{R})$, then $g(m) = x$;
*Rule 2*: If there exists a unique agent $i \in N$ such that $(\bar{R}, x) = (R^\ell, x^\ell)$ for all $\ell \in N \setminus \{i\}$ and $(R^i, x^i) \neq (\bar{R}, x)$, and $x \in F(\bar{R})$:
*Rule 2.1*: if $R^i = \bar{R}$, then $g(m) = x$;
*Rule 2.2*: if $R^i \neq \bar{R}$, then

$$g(m) = \begin{cases} x^i & \text{if } x^i \in C_i(\bar{R}, x), \\ x & \text{otherwise.} \end{cases}$$

*Rule 3*: Otherwise, $g(m) = x^{\ell^*(m)}$ where $\ell^*(m) = \sum_{i \in N} k^i \pmod{n}$.[16]

In words, the mechanism prescribes the following:

*Rule 1* applies if agents unanimously agree on a preference profile and an outcome. As a consequence, the unanimously announced outcome, $x$, is the outcome of the mechanism.

*Rule 2* applies if all agents but one (agent $i$) state the same outcome and preference profile, while agent $i$ makes a different outcome announcement or preference announcement. Then, *Rule 2.1* applies if agent $i$ disagrees with others only on the outcome announcement. In that case, the outcome of the mechanism is the outcome, $x$, announced by all other agents. On the other hand, *Rule 2.2* applies if agent $i$ announces a preference profile which differs from that announced by the others. In that case, the outcome of the mechanism is the $x^i$ announced by agent $i$, if it is an attainable outcome and not better than the outcome $x$ for $i$ when her true preference is equal to that announced by the other agents. Otherwise, the outcome is $x$.

*Rule 3* applies in all other cases and the outcome of the mechanism is determined by the agent who wins the "modulo game".

The above mechanism is a mechanism with simple punishment. Moreover, it is similar but not identical to the canonical mechanism used to prove the classical Maskin's Theorem. The difference is in the definition of *Rule 2*. While our mechanism distinguishes whether agent $i$ announces a different preference profile or not, the canonical *Rule 2* does not make this distinction.[17] Moreover, though both mechanisms satisfy the condition of simple punishment,

[15] The focus on this kind of mechanism is without loss of generality, see Maskin (1999).

[16] If the remainder is *zero*, the winner of the game is agent $n$. See Saijo (1988).

[17] In the canonical mechanism, in all cases in which all agents but one make exactly the same announcement, the outcome of the mechanism is given in the same way as in our *Rule 2.2*.

our distinction in *Rule 2* allows the planner to better exploit the fact that every partially-honest agent is making a truthful statement in equilibrium.

To explain this aspect, suppose that in equilibrium, the message profile falls into *Rule 2.1*, so that the message by agent $i$ differs from the message reported by the others only in the outcome announcement. Then, all partially-honest agents announce truthfully the preference profile; otherwise, any of the false-telling partially-honest agents can deviate to *Rule 2.2* profitably. Then, if the outcome of the mechanism is the $x$ announced by all others, we can directly conclude that $x$ is $F$-optimal at the announced preference profile. This would not be possible if the mechanism permitted the selection of $x^i \in C_i\left(\bar{R}, x\right)$, with $x^i \neq x$, announced by agent $i$.

Next, suppose that in equilibrium, the message profile falls into *Rule 2.2*. In this equilibrium, if $H \backslash \{i\} \neq \varnothing$, then every $h \in H \backslash \{i\}$ should send the truthful message because given $m \in M$ falling into *Rule 2.2*, any false-telling agent $h$ in $H \backslash \{i\}$ can always find a truth-telling $m'_h$ such that $(m'_h, m_{-h})$ corresponds to *Rule 3* with $g\left(m'_h, m_{-h}\right) = g\left(m\right)$. By a similar argument, $i$ should send the truthful message in this equilibrium, whenever $i \in H$.

We are now ready to state our second result of this sub-section; Condition $\mu^{**}$ is necessary and sufficient for partially-honest implementation when the domain of preferences is sufficiently rich and only mechanisms with simple punishment are admissible (the formal proof is relegated to Appendix).

**Theorem 2.** *Let Assumption 1 and* $\Gamma = \Gamma_{SP}$ *hold, and suppose that* $\mathcal{R}^n$ *satisfies* **RD**. *An SCC* $F \in \mathcal{F}$ *is* partially-honest implementable *if and only if it satisfies* Condition $\mu^{**}$.

## 3.2 Partially-honest implementation by $s$-mechanisms

This sub-section focuses on partially-honest implementation by $s$-mechanisms.

The basic idea behind this mechanism is to cover each agent's preference twice. For example, agent $i$'s preference may be covered by her own announcement and by that of another agent involved in the mechanism. A way to proceed is to arrange agents clockwise facing inward, and require that each agent $\ell$ announces, *inter alia*, the preferences of the agent standing immediately to her left, that is, of agent $\ell + 1$. Formally, an $s$-mechanism can be defined as follows.

**Definition 3.** A mechanism $\gamma = (M, g)$ is an $s$-mechanism if, for any $\ell \in N$, $M_\ell \equiv \mathcal{R}_\ell \times \mathcal{R}_{\ell+1} \times Y \times N$, with $n + 1 = 1$ and $Y \subseteq X$.

Thus, each agent $\ell$ announces her preference, $R^\ell_\ell$, the preference of her neighbor, $R^\ell_{\ell+1}$, an outcome, $x^\ell$, and an agent index, $k^\ell$. It is important to note that the results reported in this sub-section hold as long as each agent's preference is covered twice. It is not crucial that each agent announces her own and her neighbor's preferences.

Requiring *forthrightness* as a regularity condition, we define partially-honest implementation by $s$-mechanisms as follows.

**Definition 4.** An *SCC* $F \in \mathcal{F}$ is partially-honest implementable by an $s$-mechanism if there exists an $s$-mechanism $\gamma \equiv (M, g)$ such that:
(i) for all $R \in \mathcal{R}^n$ and all $H \in \mathcal{H}$, $F\left(R\right) = NA\left(\gamma, \succcurlyeq^R\right)$; and
(ii) for all $R \in \mathcal{R}^n$ and all $x \in F\left(R\right)$, if $m_\ell = \left(R_\ell, R_{\ell+1}, x, k^\ell\right) \in M_\ell$ for all $\ell \in N$, with $\ell + 1 = 1$ if $\ell = n$, then $m \in NE\left(\gamma, \succcurlyeq^R\right)$ and $g\left(m\right) = x$.

In Definition 4, it is required not only that all $F$-optimal outcomes coincide with partially-honest Nash equilibrium outcomes of the game $\left(\gamma, \succcurlyeq^R\right)$ defined by an $s$-mechanism - for any state $R \in \mathcal{R}^n$ and any $H \in \mathcal{H}$ -, but also that such an $s$-mechanism satisfies *forthrightness*. Forthrightness requires that if the outcome $x$ is $F$-optimal at the state $R$ and each agent announces truthfully her preference $R_\ell$ and her neighbor's preference $R_{\ell+1}$ and announces this $x$, then the message profile should be a Nash equilibrium of an $s$-mechanism and its equilibrium outcome be the announced outcome $x$.

Forthrightness was originally introduced in economic environments by Dutta *et al.* (1995) and Saijo *et al.* (1996), and it has desirable implications. A mechanism satisfying forthrightness is simple in the sense that it is easy to compute the outcome of an equilibrium message profile. Moreover, if a mechanism fails to satisfy this condition, it is subject to information smuggling; that is, the message space can be reduced to an arbitrary smaller dimensional space. Thus, any partially-honest implementable *SCC* by $s$-mechanisms would be partially-honest implementable by a 'further strategy space reduction mechanism' like *self-relevant mechanisms* (Tatamitani, 2000), unless forthrightness is required. This indicates that there is no legitimate reason for characterizing partially-honest implementation by $s$-mechanisms without forthrightness. Hence, to make sense of partially-honest implementation by this type of mechanism, we require the regularity condition of forthrightness in Definition 4.

Before turning to the findings of this sub-section, we discuss what constitutes a truthful message for $s$-mechanisms. As our objective is to examine what societal goal $F$ can be implemented when there are agents who have a minimal dishonesty aversion, we define a message of agent $\ell$ as truthful if this agent states to the planner her true preference and the true preference of her neighbor. Formally, given an $s$-mechanism $\gamma = (M, g)$, a preference profile $R \in \mathcal{R}^n$, and a societal goal $F \in \mathcal{F}$, the range of the truth-telling correspondence of agent $\ell \in N$ is

$$T_\ell^\gamma (R, F) \equiv \{(R_\ell, R_{\ell+1})\} \times Y \times N, \tag{2}$$

where $n + 1 = 1$.

The issue of what constitutes the necessary and sufficient condition for implementation by $s$-mechanisms in the conventional framework has been recently addressed by Lombardi and Yoshihara (2010), who introduce a new condition - *Condition* $M_s$ -, which is similar to Condition $M$ appearing in Sjöström (1991). This condition can be stated as follows.

Condition $M_s$ (for short, $M_s$): There exists a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there exists a profile of sets $(C_\ell (R_\ell, x))_{\ell \in N}$ such that $x \in C_\ell (R_\ell, x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; finally, for all $R^* \in \mathcal{R}^n$, the following (i)-(iii) are satisfied:
(i) if $C_\ell (R_\ell, x) \subseteq L(R_\ell^*, x)$ for all $\ell \in N$, then $x \in F(R^*)$;
(ii) for all $i \in N$, if $y \in C_i (R_i, x) \subseteq L(R_i^*, y)$ and $y \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, then $y \in F(R^*)$;
(iii) if $y \in \max_{R_\ell^*} Y$ for all $\ell \in N$, then $y \in F(R^*)$.

Notice that Condition $M_s$ differs from Condition $\mu$ only in that the set of attainable outcomes $C_\ell (R_\ell, x)$ of agent $\ell$ depends solely on her preference $R_\ell$ rather than on the entire profile $R \in \mathcal{R}^n$.

In what follows, our first task is to find necessary conditions for partially-honest implementation by $s$-mechanisms. For the same reasons highlighted in sub-section 3.1, Condition $M_s$ is too strong to constitute a necessary condition for partially-honest implementation by the type of mechanism at issue. A weaker variant of Condition $M_s$, which is relevant for our study, can be stated as follows.

Condition $M_s^*$ (for short, $M_s^*$): There exists a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there exists a profile of sets $(C_\ell(R_\ell, x))_{\ell \in N}$ such that $x \in C_\ell(R_\ell, x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; finally, for all $H \in \mathcal{H}$ and all $R^* \in \mathcal{R}^n$, the following (i)-(iii) are satisfied:
(i) if $C_\ell(R_\ell, x) \subseteq L(R_\ell^*, x)$ for all $\ell \in N$ and $x \notin F(R^*)$, then there exists $H' \subseteq H$ such that for all $h \in H'$, $(R_h, R_{h+1}) \neq (R_h^*, R_{h+1}^*)$;
(ii) for all $i \in N$, if $y \in C_i(R_i, x) \subseteq L(R_i^*, y)$, $y \in \max_{R_\ell^*} Y$ for all $\ell \in N\backslash\{i\}$, and $y \notin F(R^*)$, then there exists $H' \subseteq H$ such that:
(a) if $H' = \{i\}$, then $(y, y') \in I_i^*$ for some $y' \in C_i(R_i, x) \setminus \{y\}$;
(b) otherwise, $(R_h, R_{h+1}) \neq (R_h^*, R_{h+1}^*)$ for all $h \in H'\backslash\{i\}$;
(iii) if $y \in \max_{R_\ell^*} Y$ for all $\ell \in N$ and $y \notin F(R^*)$, then there is an $\ell \in H$ such that $(y, y') \in I_\ell^*$ for some $y' \in Y\backslash\{y\}$.

Condition $M_s^*$ stands in stark contrast to Condition $\mu^{**}$ in including a weaker variant of the Maskin monotonicity. This weakening requires that if an outcome $x$ is $F$-optimal at state $R$, and this outcome is not preferred less by any agent $\ell \in N$ than any other outcome in $C_\ell(R_\ell, x)$ at $R^*$, then $x$ must be $F$-optimal at $R^*$ whenever the preference of any potential partially-honest agent and that of her neighbor are identical between $R$ and $R^*$. In contrast, Conditions $M_s^*$(ii) and $M_s^*$(iii) are weaker versions of Conditions $M_s$(ii) and $M_s$(iii).

The next theorem shows that Condition $M_s^*$ is necessary for partially-honest implementation by $s$-mechanisms.

**Theorem 3.** *Let Assumption 1 hold. If an SCC $F \in \mathcal{F}$ is* partially-honest implementable by an $s$-mechanism, *then it satisfies* Condition $M_s^*$.

**Proof.** Let Assumption 1 hold. Let $\diamond \in N$ be an arbitrary agent index. Let $\gamma \equiv (M, g)$ be an $s$-mechanism which partially-honest implements $F \in \mathcal{F}$. Let $Y \equiv g(M)$. Take any $H \in \mathcal{H}$, any $R \in \mathcal{R}^n$, and any $x \in F(R)$. For all $\ell \in N$, let $C_\ell(R_\ell, x) \equiv g(M_\ell, m_{-\ell}(R, x))$ where $m_{-\ell}(R, x)$ is such that $m_i(R, x) = (R_i, R_{i+1}, x, \diamond) \in M_i$ for all $i \in N\backslash\{\ell\}$, with $n + 1 = 1$. By forthrightness, $m(R, x) = (m_\ell(R, x), m_{-\ell}(R, x)) \in NE(\gamma, \succsim^R)$ holds for all $H' \in \mathcal{H}$, and $g(m(R, x)) = x$. Then, $C_\ell(R_\ell, x) = g(M_\ell, m_{-\ell}(R, x)) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$. We show that $F$ satisfies Conditions $M_s^*$(i)-$M_s^*$(iii). As it is easy to see that $F$ meets $M_s^*$(iii), we omit its proof here. Take any $H \in \mathcal{H}$ and any $R^* \in \mathcal{R}^n$.

Suppose that $C_\ell(R_\ell, x) \subseteq L(R_\ell^*, x)$ for all $\ell \in N$ and $x \notin F(R^*)$. Then, since $C_\ell(R_\ell, x) = g(M_\ell, m_{-\ell}(R, x))$ for all $\ell \in N$, it follows that for each $H \in \mathcal{H}$, there exists an $H' \subseteq H$ such that for all $h \in H'$, $m_h(R, x) \notin T_h^\gamma(R^*, F)$ and $(g(m_h', m_{-h}(R, x)), g(m(R, x))) \in I_h^*$ for some $m_h' \in T_h^\gamma(R^*, F)$. Thus, $(R_h^*, R_{h+1}^*) \neq (R_h, R_{h+1})$ for all $h \in H'$. Hence, $F$ satisfies Condition $M_s^*$(i).

Pick any $i \in N$. Suppose that $y \in C_i(R_i, x) \subseteq L(R_i^*, y)$, $y \in \max_{R_\ell^*} Y$ for all $\ell \in N\backslash\{i\}$, and $y \notin F(R^*)$. Then, since $C_i(R_i, x) = g(M_i, m_{-i}(R, x))$, $g(m_i, m_{-i}(R, x)) = y$ for some $m_i \in M_i$. Let $\widehat{m} \equiv (m_i, m_{-i}(R, x))$. Moreover, as $y \notin F(R^*) = NA(\gamma, \succsim^{R^*})$ for all $\overline{H} \in \mathcal{H}$, it follows that for each $\overline{H} \in \mathcal{H}$, there exists an $H' \subseteq \overline{H}$ such that for all $h \in H'$, $\widehat{m}_h \notin T_h^\gamma(R^*, F)$ and $(g(m_h^*, \widehat{m}_{-h}), g(\widehat{m})) \in I_h^*$ for some $m_h^* \in T_h^\gamma(R^*, F)$. Let $H' = \{i\} \subseteq H$ for the given $H \in \mathcal{H}$, and $\{y\} = \max_{R_i^*} C_i(R_i, x)$. It follows that $g(m_i^*, m_{-i}) = y$ which leads to $(m_i^*, m_{-i}) \in NE(\gamma, \succsim^{R^*})$ for this $H$, a contradiction. Thus, $F$ satisfies $M_s^*$(ii.a). Finally, let $H' \neq \{i\}$ for $H' \subseteq H$. It can readily be obtained by the definition of $H'$ that $F$ satisfies $M_s^*$(ii.b). ■

A slight strengthening of Condition $M_s^*$ is required for the sufficiency result. The two auxiliary conditions which are required are the standard Condition $\mu$(iii) - or equivalently, Condition $M_s$(iii) - and Condition $\mu^{**}$(iv). This condition can be stated as follows.

Condition $M_s^{**}$ (for short, $M_s^{**}$): There exists a set $Y \subseteq X$ and, for all $H \in \mathcal{H}$, all $R \in \mathcal{R}^n$, and all $x \in F(R)$, there exists a profile of sets $(C_\ell(R_\ell, x))_{\ell \in N}$ such that $x \in C_\ell(R_\ell, x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; Condition $M_s^*$ and Condition $M_s$(iii) hold; finally, for all $H \in \mathcal{H}$ and all $R^* \in \mathcal{R}^n$, Condition $\mu^{**}$(iv) holds.[18]

The above condition is not only sufficient when the domain of preferences is rich enough, but also necessary when only $s$-mechanisms with simple punishments are admissible. Before stating this result (whose proof is relegated to Appendix), it may be worthwhile describing the mechanism constructed to obtain the sufficiency part.

The implementing mechanism uses the idea of cyclic announcement of messages proposed in Saijo (1988), and is identical to the $s$-mechanism used to prove that Condition $M_s$ is necessary and sufficient for implementation by $s$-mechanisms in the conventional framework (Lombardi and Yoshihara, 2010). In line with Lombardi and Yoshihara (2010), for an $s$-mechanism $\gamma = (M, g)$, we say that the message profile $m \in M$ is:

(i) *consistent* with $R$ and $x$ if, for all $j \in N$, $R_j^j = R_j^{j-1} = R_j$ and $x^j = x$;
(ii) $m_{-i}$ *quasi-consistent* with $R$ and $x$, where $i \in N$, if for all $j \in N$, $x^j = x$, and for all $j \in N \backslash \{i, i+1\}$, $R_j^j = R_j^{j-1} = R_j$, $R_i^{i-1} = R_i$, $R_{i+1}^{i+1} = R_{i+1}$, and $[R_i^i \neq R_i$ or $R_{i+1}^i \neq R_{i+1}]$;
(iii) $m_{-i}$ *consistent* with $R$ and $x$, where $i \in N$, if for all $j \in N \backslash \{i\}$, $x^j = x \neq x^i$, and for all $j \in N \backslash \{i, i+1\}$, $R_j^j = R_j^{j-1} = R_j$, $R_i^{i-1} = R_i$, $R_{i+1}^{i+1} = R_{i+1}$;

where $1 - 1 = n$.

In words, a message profile is consistent with an outcome $x$ and a preference profile $R$ if there is no break in the cyclic announcement of preferences and all agents announce the outcome $x$. On the other hand, it is quasi-consistent with $x$ and $R$ if there are at most two consecutive breaks in the cyclic announcement of preferences, and $x$ is unanimously announced. Finally, a message profile $m$ is $m_{-i}$ consistent with $x$ and $R$ if agent $i$ announces an outcome different from the outcome $x$ announced by the others, if there are no more than two consecutive breaks in the cyclic announcement of preferences, and, finally, if these breaks (if any) happen only in the announcement of the preference made by agent $i$.

Define the outcome function $g$ as follows. For any message profile $m \in M$,

*Rule 1*: If $m$ is consistent with $(\bar{R}, x) \in \mathcal{R}^n \times Y$ and $x \in F(\bar{R})$, then $g(m) = x$.
*Rule 2*: If for some $i \in N$, $m$ is $m_{-i}$ is quasi-consistent with $(\bar{R}, x) \in \mathcal{R}^n \times Y$ and $x \in F(\bar{R})$, then $g(m) = x$.
*Rule 3*: If for some $i \in N$, $m$ is $m_{-i}$ consistent with $(\bar{R}, x) \in \mathcal{R}^n \times Y$, $x \in F(\bar{R})$, and $C_i(\bar{R}_i, x) \neq Y$, then

$$g(m) = \begin{cases} x^i & \text{if } \quad x^i \in C_i(\bar{R}_i, x) \\ x & \qquad \text{otherwise.} \end{cases}$$

*Rule 4*: Otherwise, $g(m) = x^{\ell^*(m)}$ where $\ell^*(m) \equiv \sum_{i \in N} k^i \pmod{n}$.

The above mechanism is one with simple punishment.

Before closing this sub-section, it may be worthwhile to provide the reason for why Condition $M_s^*$(i) is required to guarantee partially-honest implementation by $s$-mechanisms. To this end, let $R$ be the true state of the world and $m$ be an Nash equilibrium message

[18] Henceforth, Condition $M_s$(iii) and Condition $\mu^{**}$(iv) are referred to as Condition $M_s^{**}$(iii) and Condition $M_s^{**}$(iv), respectively. Moreover, we refer to the statement that requires only one of the statements (i) and (ii) in Condition $M_s^*$ as Conditions $M_s^{**}$(i) and $M_s^{**}$(ii).

profile of the game $\left(\gamma, \succcurlyeq^{R}\right)$ which falls into *Rule 1.* When a canonical mechanism is employed and an equilibrium message profile falls into *Rule 1* - of the mechanism described in the previous sub-section -, the preference profile $R^{i}$ is announced truthfully; that is, $R^{i} = R$, and this permitted us to conclude in Theorem 2 that the unanimously announced outcome was $F$-optimal at $R$. This conclusion, however, is no longer possible when we are dealing with $s$-mechanisms. The reason is that even though all partially-honest agents are reporting truthfully, it is in general not possible to reconstruct the true state $R$ from their reports. Therefore, Condition $M_{s}^{**}$(i) is required to guarantee that $x$ is $F$-optimal at $R$.

To conclude, the following theorem shows that Condition $M_{s}^{**}$ is necessary and sufficient for partially-honest implementation by $s$-mechanisms under the same mild requirements stated in Theorem 2.

**Theorem 4.** *Let Assumption 1 and* $\Gamma = \Gamma_{SP}$ *hold, and let* $\mathcal{R}^{n}$ *satisfy* **RD**. *An SCC* $F \in \mathcal{F}$ *is* partially-honest implementable by an $s$-mechanism *if and only if* $F$ *satisfies* Condition $M_{s}^{**}$.

## 3.3 Partially-honest implementation by self-relevant mechanisms

From the viewpoint of informational decentralization in mechanisms, it is desirable that an agent discloses information related only to her own characteristics (Hurwicz, 1960, 1972). This entails analyzing partially-honest implementation by *self-relevant mechanisms*, which is the subject of this sub-section.

The type of self-relevant mechanism considered herein is that introduced in Tatamitani (2001). In this mechanism, each participanting agent $\ell$ is required to state an outcome, $x^{\ell}$, an agent index, $k^{\ell}$, and only her preference, $R_{\ell}^{\ell}$. Formally, it can be defined as follows.

**Definition 5.** A mechanism $\gamma \equiv (M, g)$ is a self-relevant mechanism if, for any $\ell \in N$, $M_{\ell} \equiv \mathcal{R}_{\ell} \times Y \times N$.

Similar to the definition of truthfulness of the previous sub-sections, a message from each agent in a self-relevant mechanism is truthful if it conveys to the planner her true preference. Formally, given a self-relevant mechanism $\gamma = (M, g)$, a preference profile $R \in \mathcal{R}^{n}$, and a societal goal $F \in \mathcal{F}$, the range of the truth-telling correspondence of agent $\ell \in N$ is

$$T_{\ell}^{\gamma}(R, F) \equiv \{R_{\ell}\} \times Y \times N. \tag{3}$$

Given this definition of truth-telling correspondences, partially-honest implementation by self-relevant mechanisms is defined as follows:

**Definition 6.** An *SCC* $F \in \mathcal{F}$ is partially-honest implementable by a self-relevant mechanism if there exists a self-relevant mechanism $\gamma \equiv (M, g)$ such that:
(i) for all $R \in \mathcal{R}^{n}$ and all $H \in \mathcal{H}$, $F(R) = NA\left(\gamma, \succcurlyeq^{R}\right)$; and
(ii) for all $R \in \mathcal{R}^{n}$ and all $x \in F(R)$, if $m_{\ell} = \left(R_{\ell}, x, k^{\ell}\right) \in M_{\ell}$ for all $\ell \in N$, then $m \in NE\left(\gamma, \succcurlyeq^{R}\right)$ and $g(m) = x$.

Like Tatamitani (2001), the above definition requires the regularity condition of forthrightness to avoid the problem of information smuggling. Moreover, Definition 6 requires that the planner has to design a self-relevant mechanism where only the $F$-optimal outcomes are realized as equilibrium outcomes of the devised game form, regardless of the current state $R$ and who is partially honest.

Let us begin with stating the necessary and sufficient conditions for implementation by self-relevant mechanisms in the conventional framework. To this end, additional notation is needed. For any $\ell \in N$, any $R_{-\ell} \in \mathcal{R}^{n-1}$, and $x \in X$, let $F_\ell^{-1}(R_{-\ell}, x) \equiv \{R'_\ell \in \mathcal{R}_\ell | x \in F(R'_\ell, R_{-\ell})\}$ and $\Lambda_\ell^F(R_{-\ell}, x) \equiv \cap_{R_\ell \in F_\ell^{-1}(R_{-\ell},x)} L(R_\ell, x)$. Given $(R, x) \in \mathcal{R}^n \times X$, define $D(R, x) \equiv \left\{\ell \in N | F_\ell^{-1}(R_{-\ell}, x) \neq \varnothing\right\}$.

The necessary and sufficient condition devised by Tatamitani (2001) can be stated as follows.

Condition $\lambda$ (for short, $\lambda$): There exists a set $Y \subseteq X$ and, for all $(R, x) \in \mathcal{R}^n \times X$ with $D(R, x) \neq \varnothing$, there exists a profile of sets $(C_\ell(R_{-\ell}, x))_{\ell \in N}$ such that $x \in C_\ell(R_{-\ell}, x) \subseteq \Lambda_\ell^F(R_{-\ell}, x) \cap Y$ for all $\ell \in D(R, x)$; finally, for all $R^* \in \mathcal{R}^n$, the following (i)-(iv) are satisfied:
(i) if $x \in F(R)$ and $C_\ell(R_{-\ell}, x) \subseteq L(R^*_\ell, x)$ for all $\ell \in N$, then $x \in F(R^*)$;
(ii) for all $i \in D(R, x)$, if $y \in C_i(R_{-i}, x) \subseteq L(R^*_i, y)$ and $y \in \max_{R^*_\ell} Y$ for all $\ell \in N \backslash \{i\}$, then $y \in F(R^*)$;
(iii) if $y \in \max_{R^*_\ell} Y$ for all $\ell \in N$, then $y \in F(R^*)$;
(iv) there exists an outcome $p(R, x) \in X$ such that:
(a) $p(R, x) \in C_\ell(R_{-\ell}, x)$ for all $\ell \in D(R, x)$;
(b) if $C_i(R_{-i}, x) \subseteq L(R^*_i, p(R, x))$ for all $i \in D(R, x)$ and $p(R, x) \in \max_{R^*_\ell} Y$ for all $\ell \in N \backslash D(R, x)$, then $p(R, x) \in F(R^*)$.

Condition $\lambda$ is markedly stronger than Condition $\mu$. Notable parts of Condition $\lambda$ are Condition $\lambda$(i) and Condition $\lambda$(iv). Condition $\lambda$(i) is much stronger than Maskin monotonicity.[19]

For the same reasons highlighted in sub-section 3.1, the participation of dishonest averse agents in a self-relevant mechanism makes Condition $\lambda$ no longer a necessary condition for implementation. A weakening of Condition $\lambda$ which is relevant for the present study can be stated as follows.

Condition $\lambda^*$ (for short, $\lambda^*$): There exists a set $Y \subseteq X$ and, for all $(R, x) \in \mathcal{R}^n \times X$ with $D(R, x) \neq \varnothing$, there exists a profile of sets $(C_\ell(R_{-\ell}, x))_{\ell \in N}$ such that $x \in C_\ell(R_{-\ell}, x) \subseteq \Lambda_\ell^F(R_{-\ell}, x) \cap Y$ for all $\ell \in D(R, x)$; finally, for all $H \in \mathcal{H}$ and all $R^* \in \mathcal{R}^n$, the following (i)-(iv) are satisfied:
(i) if $x \in F(R)$, $C_\ell(R_{-\ell}, x) \subseteq L(R^*_\ell, x)$ for all $\ell \in N$, and $x \notin F(R^*)$, then there exists $H' \subseteq H$ such that for all $h \in H'$, $R_h \neq R^*_h$;
(ii) for all $i \in D(R, x)$, if $y \in C_i(R_{-i}, x) \subseteq L(R^*_i, y)$, $y \in \max_{R^*_\ell} Y$ for all $\ell \in N \backslash \{i\}$, and $y \notin F(R^*)$, then there exists $H' \subseteq H$ such that:
(a) if $H' = \{i\}$, then $(y, y') \in I^*_i$ for some $y' \in C_i(R_{-i}, x) \backslash \{y\}$;
(b) otherwise, $R_h \neq R^*_h$ for all $h \in H' \backslash \{i\}$;
(iii) if $y \in \max_{R^*_\ell} Y$ for all $\ell \in N$ and $y \notin F(R^*)$, then there exists an $\ell \in H$ such that $(y, y') \in I^*_\ell$ for some $y' \in Y \backslash \{y\}$;
(iv) there exists an outcome $p(R, x) \in X$ such that:
(a) $p(R, x) \in C_\ell(R_{-\ell}, x)$ for all $\ell \in D(R, x)$;
(b) if $C_i(R_{-i}, x) \subseteq L(R^*_i, p(R, x))$ for all $i \in D(R, x) \neq \varnothing$, $p(R, x) \in \max_{R^*_\ell} Y$ for all $\ell \in N \backslash D(R, x)$, and $p(R, x) \notin F(R^*)$, then $R_h \neq R^*_h$ for some $h \in H$.

It is important to note that Condition $\lambda^*$ incorporates not only a monotonicity-type condition but also a punishment-type condition, Condition $\lambda^*$(iv).

[19] See Tatamitani (2002) for a detailed analysis on how restrictive Condition $\lambda$(i) is.

Though Condition $\lambda^*$ is weaker than Condition $\lambda$, these two conditions are very similar. Nonetheless, the next theorem shows that the proposed amendment of Condition $\lambda$ is a necessary condition for partially-honest implementation by self-relevant mechanisms.

**Theorem 5.** *Let Assumption 1 hold. If an SCC $F \in \mathcal{F}$ is* partially-honest implementable by *a* self-relevant mechanism, *then it satisfies* Condition $\lambda^*$.

**Proof.** Let Assumption 1 hold and $\diamond \in N$ be an arbitrary agent index. Let $\gamma \equiv (M, g)$ be a self-relevant mechanism which partially-honest implements $F \in \mathcal{F}$. Let $Y \equiv g(M)$. Take any $H \in \mathcal{H}$ and any $(R, x) \in \mathcal{R}^n \times X$ with $D(R, x) \neq \varnothing$. For any $i \in D(R, x)$, let $C_i(R_{-i}, x) \equiv g(M_i, m_{-i})$, where $m_{-i}$ is such that $m_\ell = (R_\ell, x, \diamond) \in M_\ell$ for all $\ell \in N \backslash \{i\}$. Therefore, $C_i(R_{-i}, x) \subseteq Y$. Next, to show $x \in C_i(R_{-i}, x) \subseteq \Lambda_i^F(R_{-i}, x)$, take any $R_i' \in F_i^{-1}(R_{-i}, x)$ and let $m_i' = (R_i', x, \diamond) \in M_i$. By forthrightness, $(m_i', m_{-i}) \in NE\left(\gamma, \succcurlyeq^{(R_i', R_{-i})}\right)$ for any $H \in \mathcal{H}$, and $g(m_i', m_{-i}) = x$. So, $x \in g(M_i, m_{-i}) \subseteq L(R_i', x)$. Since it holds for any $R_i' \in F_i^{-1}(R_{-i}, x)$, we have that $x \in C_i(R_{-i}, x) \subseteq \Lambda_i^F(R_{-i}, x)$. Note that the proofs that $F$ meets Conditions $\lambda^*$(i)-$\lambda^*$(iii) can be obtained by following a reasoning similar to that used in Theorem 3, so we omit them here. Finally, we show that $F$ satisfies Condition $\lambda^*$(iv).

Take any $(R, x) \in \mathcal{R}^n \times X$ with $D(R, x) \neq \varnothing$. Let $m \in M$ be such that $m_\ell = (R_\ell, x, \diamond) \in M_\ell$ for all $\ell \in N$. Let $g(m) \equiv p(R, x)$. Then, $p(R, x) \in g(M_\ell, m_{-\ell}) \subseteq Y$ for all $\ell \in D(R, x)$. Furthermore, suppose that $C_i(R_{-i}, x) \subseteq L(R_i^*, p(R, x))$ for all $i \in D(R, x)$, $p(R, x) \in \max_{R_\ell^*} Y$ for all $\ell \in N \backslash D(R, x)$, and $p(R, x) \notin F(R^*)$. Then, by partially-honest implementability of $F$ by $\gamma$, it follows that $m \notin NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for any $\overline{H} \in \mathcal{H}$. Then, given $H \in \mathcal{H}$, $m_h \notin T_h^\gamma(R^*, F)$ for some $h \in H$. Thus, $F$ satisfies Condition $\lambda^*$(iv). ■

Condition $\lambda^*$ alone does not guarantee the sufficiency of partially-honest implementation by self-relevant mechanisms. As a necessary and sufficient condition, it needs a slight strengthening of Condition $\lambda^*$ by adding the two auxiliary conditions, Condition $\mu$(iii) and Condition $\mu^{**}$(iv). Thus, the new condition as a whole is stated below.

Condition $\lambda^{**}$ (for short, $\lambda^{**}$): There exists a set $Y \subseteq X$ and, for all $(R, x) \in \mathcal{R}^n \times X$ with $D(R, x) \neq \varnothing$, there exists a profile of sets $(C_\ell(R_{-\ell}, x))_{\ell \in N}$ such that $x \in C_\ell(R_{-\ell}, x) \subseteq \Lambda_\ell^F(R_{-\ell}, x) \cap Y$ for all $\ell \in D(R, x)$; Condition $\lambda^*$ and Condition $\mu$(iii) hold; finally, for all $H \in \mathcal{H}$ and all $R^* \in \mathcal{R}^n$, Condition $\mu^{**}$(iv) holds.[20]

**Theorem 6.** *Let Assumption 1 and $\Gamma = \Gamma_{SP}$ hold, and let $\mathcal{R}^n$ satisfy* **RD**. *An SCC $F \in \mathcal{F}$ is* partially-honest implementable by a self-relevant mechanism *if and only if it satisfies* Condition $\lambda^{**}$.

# 4 Implications

This section briefly discusses the implications of the results reported in section 3.

Before going into the details, let us note that we cannot specify in advance the strucutre of the set $\mathcal{H}$ in which the analysis takes place. By our assumption, $\mathcal{H}$ could be anything whenever $\mathcal{H} \subseteq 2^N \backslash \varnothing$ and $\#\mathcal{H} \geqq 2$ hold. However, when we examine the performance of each *SCC* in terms of its partially-honest implementability, it seems most plausible to proceed with this examination by assuming $\mathcal{H} = 2^N \backslash \varnothing$. This is because such an assumption

[20] Henceforth, Condition $\mu$(iii) and Condition $\mu^{**}$(iv) are referred to as Condition $\lambda^{**}$(iii) and Condition $\lambda^{**}$(iv), respectively. Moreover, we refer to the statement that requires only one of the statements (i)-(ii) in Condition $\lambda^*$ as Conditions $\lambda^{**}$(i)-$\lambda^{**}$(ii).

implies the severest situation for the planner in the sense that she cannot know even the class of potential sets of partially-honest agents, and so she cannot help but simply presume $\mathcal{H} = 2^N \backslash \varnothing$, and then design a mechanism which can implement her goal, $F$, even in this situation. Indeed, by covering the case that $\mathcal{H} = 2^N \backslash \varnothing$, the planner is ensured of the implementability of $F$ for any other specialization that the set $\mathcal{H}$ may take. For this reason, we turn to analyze some implications of the aforementioned theorems under the specification that the structure of $\mathcal{H}$ is $\mathcal{H} = 2^N \backslash \varnothing$.

The first proposition is an impossibility, showing that Condition $\mu^*$ imposes non-trivial restrictions on the class of partially-honest implementable *SCC*s. To show this result, let us define the *Pareto SCC*. For each $R \in \mathcal{R}^n$, the *Pareto set*, $PO(R)$, is:[21]

$$PO(R) \equiv \{x \in X | \nexists y \in X \colon \ (y, x) \in R_i \text{ for all } i \in N \text{ and } \ (y, x) \in P_i \text{ for some } i \in N\}.$$

An *SCC* $F$ on $\mathcal{R}^n$ is the *Pareto SCC*, denote $F^{PO}$, if $F(R) = PO(R)$ for all $R \in \mathcal{R}^n$. Our next result shows that this *SCC* violates Condition $\mu^*$.

**Proposition 1.** *Let* Assumption 1 *hold.* $F^{PO}$ *on* $\mathcal{R}^n$ *is not* partially-honest implementable *if* $\mathcal{H} = 2^N \backslash \varnothing$.

**Proof.** Assume, to the contrary, that $F^{PO}$ satisfies Condition $\mu^{**}$. Let $N = \{1, 2, 3\}$ with $\#N = 3$, $X = \{x, y, z\}$ with $\#X = 3$, and $\mathcal{R}^3 = \{R, R^*\}$, where agents' preferences are as follows:

| $R$ | | | $R^*$ | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 2 | 3 |
| $x$ | $y$ | $z$ | $x$ | $x, y$ | $x, y$ |
| $y$ | $z$ | $x$ | $y$ | $z$ | $z$ |
| $z$ | $x$ | $y$ | $z$ | | |

where, as usual, $\substack{x \\ y}$ means that the agent in question strictly prefers $x$ to $y$, while $x, y$ means that the agent at issue is indifferent between $x$ and $y$.

As $y \in PO(R)$, there exists a profile $(C_\ell(R, y))_{\ell \in N}$ such that $y \in C_\ell(R, y) \subseteq L(R_\ell, y) \cap Y$ for all $\ell \in N$. Since $PO(R) = X$, it follows that $Y = X$. Notice that Condition $\mu^*$(ii.a) is vacuously satisfied if $H = \{i\} \subseteq \{2, 3\}$. Then, let $H = \{1\}$. Observe $y \in \max_{R^*_\ell} X$ for all $\ell \in \{2, 3\}$ and $y \in C_1(R, y) \subseteq L(R_1, y) = L(R^*_1, y)$. Condition $\mu^*$(ii.a) implies that $y \in F^{PO}(R^*) \neq PO(R^*) = \{x\}$, a contradiction. ■

The next proposition is a possibility result, showing that while the *Pareto SCC*, $F^{PO}$, defined on the domain of *single-plateaued preferences* violates both Condition $\mu$(i) and Condition $\mu$(ii), it is partially-honest implementable by virtue of Theorem 2. Before proving this result, let us define the environment in which the result is formulated.

Let $M \in \mathbb{R}_{++}$ be an amount of some infinitely divisible commodity which has to be allocated among a set of agents $N$, with $n \geq 3$. An allocation is a list $x \in \mathbb{R}^n_+$ such that $\sum x_\ell = M$.[22] Let $X \equiv \left\{x \in \mathbb{R}^n_+ | \sum x_\ell = M\right\}$ be the set of feasible allocations. Each agent $\ell \in N$ is equipped with a continuous and single-plateaued preference relation $R_\ell$ defined on $X$ as follows: there exists a continuous and quasi-concave real-valued function $u_{R_\ell} : [0, M] \to \mathbb{R}$ such that, for any $x, x' \in X$, $u_{R_\ell}(x_\ell) \geq u_{R_\ell}(x'_\ell) \Leftrightarrow (x, x') \in R_\ell$. For each $\ell \in N$, the preference relation $R_\ell$ defined on $X$ is called *single-plateaued* when there exist

[21] Henceforth, the symbol $\nexists$ denotes the negation of the the existence quantifier, $\exists$.

[22] When its bounds are not explicitly indicated, a summation should be understood to cover all agents.

two numbers $\bar{x}_\ell, \underline{x}_\ell \in [0, M]$ such that $\underline{x}_\ell \leq \bar{x}_\ell$ and for all $x_\ell, y_\ell \in [0, M]$: (i) if $x_\ell < y_\ell \leq \underline{x}_\ell$ or $x_\ell > y_\ell \geq \bar{x}_\ell$, then $(y', x') \in P_\ell$ for any $x', y' \in X$, with $x'_\ell = x_\ell$ and $y'_\ell = y_\ell$; (ii) if $x_\ell, y_\ell \in [\underline{x}_\ell, \bar{x}_\ell]$, then $(x', y') \in I_\ell$ for any $x', y' \in X$, with $x'_\ell = x_\ell$ and $y'_\ell = y_\ell$. The interval $p(R_\ell) \equiv [\underline{x}_\ell, \bar{x}_\ell]$ is the *plateau* of $R_\ell$, where $\underline{x}$ is the left end-point of the plateau of $R_\ell$, and $\bar{x}$ is the right end-point. Let $\bar{\mathcal{R}}_\ell$ be the class of all such preference relations for each agent $\ell \in N$. Note that by definition of $R_\ell \in \bar{\mathcal{R}}_\ell$, it follows that $R_\ell$ is single-peaked if $\underline{x}_\ell = \bar{x}_\ell$. Given $x_\ell \in [0, M]$, let $r_\ell(x_\ell)$ be the consumption bundle on the other side of agent $\ell$'s plateau amounts that she finds indifferent to $x_\ell$ if such consumption exists, and the end-point of $[0, M]$ on the other side of her plateau amounts otherwise. Given a profile of preferences $R \in \bar{\mathcal{R}}^n$, $p(R) \equiv (p(R_1), ..., p(R_n))$ denotes its associated profile of plateau amounts.

We are now in a position to establish our possibility result.

**Proposition 2.** *Let $F^{PO}$ on $\bar{\mathcal{R}}^n$ be the* Pareto *SCC. Then, (i) $F^{PO}$ satisfies neither of* Conditions $\mu$(i) and $\mu$(ii)*; (ii) given Assumption 1, $F^{PO}$ satisfies* Condition $\mu^{**}$.

**Proof.** Let $F^{PO}$ on $\bar{\mathcal{R}}^n$ be the *Pareto SCC*.

We illustrate part (i) by considering the following the three-agent example.[23]

Let $M = 1$, $N \equiv \{1, 2, 3\}$, with $\#N = 3$, and $R, R^* \in \bar{\mathcal{R}}^n$ be such that $R_1 = R_1^*$, $p(R) = \left(\frac{1}{4}, 1, [0, 1]\right)$, and $p(R^*) = \left(\frac{1}{4}, \left[\frac{1}{2}, 1\right], [0, 1]\right)$. Let $x = \left(\frac{1}{6}, \frac{5}{6}, 0\right)$ and $y = \left(\frac{1}{5}, \frac{4}{5}, 0\right)$. First note that $x, y \in X$, $x \in PO(R)$, and $L(R_1, x) = L(R_1^*, x) = \left\{z \in X \mid 0 \leq z_1 \leq \frac{1}{6} \text{ or } r_1(x_1) \leq z_1 \leq 1\right\}$, $L(R_2, x) = \left\{z \in X \mid 0 \leq z_2 \leq \frac{5}{6}\right\}$ and $L(R_3, x) = L(R_3^*, x) = L(R_2^*, x) = X$. Moreover, note that $y \notin L(R_1, x)$ while $y \in L(R_2, x)$. Suppose that $F^{PO}$ satisfies Conditions $\mu$(i) and $\mu$(ii). Note that $x, y \in \max_{R_2^*} X \cap \max_{R_3^*} X$. Furthermore, for any $C_1(R, x) \subseteq L(R_1, x)$, it follows that $C_1(R, x) \subseteq L(R_1^*, x)$. Condition $\mu$(ii) implies that $x \in F^{PO}(R^*)$. However, $x \notin PO(R^*)$ since $y$ Pareto dominates it, a contradiction. Also, since $x \in F^{PO}(R)$ and $L(R_\ell, x) \subseteq L(R_\ell^*, x)$ for all $\ell \in N$, Condition $\mu$(i) implies that $x \in F^{PO}(R^*)$, a contradiction.

To show part (ii), let $(R, x, \ell) \in \bar{\mathcal{R}}^n \times X \times N$ with $x \in F^{PO}(R)$, and let $C_\ell(R, x) \equiv L(R_\ell, x)$. Also, $X = Y$ as $F^{PO}$ satisfies unanimity. We will show that $F^{PO}$ satisfies Condition $\mu^{**}$ under these specifications. Pick any arbitrary $(R, R^*, x) \in \bar{\mathcal{R}}^n \times \bar{\mathcal{R}}^n \times X$, with $x \in F^{PO}(R)$. Condition $\mu^{**}$(i) is always vacuously satisfied. Moreover, $F^{PO}$ meets Condition $\mu^{**}$(iii). Next, we show that $F^{PO}$ satisfies $\mu^{**}$(ii) and $\mu^{**}$(iv).

Take any $(H, i) \in \mathcal{H} \times N$. Suppose that $y \in C_i(R, x) = L(R_i, x) \subseteq L(R_i^*, y)$ and $y \in \max_{R_\ell^*} X$ for all $\ell \in N \backslash \{i\}$.

Let $H = \{i\}$ and $y \notin F^{PO}(R^*)$. We show that $\{y\} \neq \max_{R_i^*} C_i(R, x)$. As $y \notin F^{PO}(R^*)$, it follows that there exists an allocation $z \in X$ such that $(z, y) \in R_j^*$ for all $j \in N$ and $(z, y) \in P_j^*$ for some $j \in N$. As $y \in \max_{R_\ell^*} X$ for all $\ell \in N \backslash \{i\}$, it follows that $(z, y) \in P_i^*$ and $(z, y) \in I_\ell^*$ for all $\ell \in N \backslash \{i\}$; moreover, $z \notin L(R_i^*, y) \supseteq L(R_i, x)$ as $(z, x) \in P_i^*$. Then, $y$ is not a plateau amount for agent $i$, and so $L(R_i^*, y) \neq X$. Let $y' \equiv (y_i, w_{-i}) \neq y$ where $w_{-i} \in \mathbb{R}_+^{n-1}$ such that $\sum_{\ell \in N \backslash \{i\}} w_\ell = \sum_{\ell \in N \backslash \{i\}} y_\ell$. The allocation $y'$ exists and belongs to the set $L(R_i, x)$ as $(x, y) \in R_i$ and $(y, y') \in I_i$. As $y' \in L(R_i, x) \setminus \{y\}$ and $(y, y') \in I_i^*$, we have that $\{y\} \neq \max_{R_i^*} C_i(R, x)$. Hence, $F^{PO}$ satisfies Condition $\mu^{**}$(ii.a).

Let $i \in H$ and $\#H > 1$. Assume that $R^* = R$ and $\{y\} = \max_{R_i^*} C_i(R, x)$. Thus, $x = y$. and so $y \in F^{PO}(R^*)$. Therefore, $F^{PO}$ satisfies Condition $\mu^{**}$(ii.b).

[23] The *Pareto SCC* is monotonic and satisfies no-veto power when $\bar{\mathcal{R}}^n$ consists only of single-peaked preference profiles.

Let $i \notin H$ and $R^* = R$. It follows that $(y,x) \in I_i$ and $(y,x) \in I_\ell$ for all $\ell \in N\setminus\{i\}$. Suppose that $y \notin F^{PO}(R)$. Then, there exists a $z \in X$ such that $(z,y) \in R_k$ for all $k \in N$ and $(z,y) \in P_j$ for some $j \in N$. By transitivity of $R_j$ for all $j \in N$, it follows that $z$ Pareto dominates $x$ under the state $R$. Then, $x \notin F^{PO}(R)$, a contradiction. Therefore, $F^{PO}$ satisfies Condition $\mu^{**}$(ii.c).

Let $H = \{i\}$, $x = y$, $R_{-i} = R^*_{-i}$, and $L(R_i, x) = L(R^*_i, x)$. We show that $x \in F^{PO}(R^*)$. Assume, to the contrary, that $x \notin F^{PO}(R^*)$. Then, there exists an allocation $z \in X$ such that $(z,x) \in R^*_k$ for all $k \in N$ and $(z,x) \in P^*_j$ for some $j \in N$. As $x \in \max_{R^*_\ell} X$ for all $\ell \in N\setminus\{i\}$, it follows that $(z,x) \in P^*_i$ and $(z,x) \in I^*_\ell$ for all $\ell \in N\setminus\{i\}$; and $(z,x) \in I_\ell$ for all $\ell \in N\setminus\{i\}$ as $R_{-i} = R^*_{-i}$. Thus, $z \notin L(R^*_i, x) = L(R_i, x)$ as $(z,x) \in P^*_i$. It follows that $x \notin F^{PO}(R)$, a contradiction. Hence, $F^{PO}$ satisfies $\mu^{**}$(iv). ■

In their seminal paper, Dutta and Sen (2009) showed that the only requirement of *no-veto power* is sufficient for partially-honest implementation. The above finding shows that the scope of implementation is further enlarged to include many *SCC*s which are non-monotonic and violate the auxiliary condition of no-veto power.

The last objective of this section is to investigate how the monotonicity-type condition incorporated in Condition $M^{**}_s$ affects partially-honest implementability. The analysis reveals that this condition is restrictive, though it is weaker than Maskin monotonicity. Remarkably, it shows that the equivalent relationship between implementation and implementation by $s$-mechanisms holding in the classical implementation framework no longer holds when there exist agents who are dishonest averse.

To this end, let us turn to define the environment in which the analysis is carried out. Let $X$ be a finite set of outcomes. For any $x, y \in X$, with $x \neq y$, and $R \in \mathcal{P}^n$, let $N_R(x,y) \equiv \{i \in N | (x,y) \in R_i\}$.[24] Let us denote $(x,y) \in T_R$ if and only if $\#N_R(x,y) \geq \#N_R(y,x)$, which implies that $x$ is majority preferred to $y$ at the profile $R$. For the sake of simplicity, suppose that $n$ is an odd number so that the majority relation $T_R$ on $X$ is a *tournament* for any $R \in \mathcal{P}^n$.[25] The set of all top-cycle outcomes at state $R \in \mathcal{P}^n$ can be defined as follows:

$$x \in TC(R) \Leftrightarrow \begin{array}{l} \forall y \in X\setminus\{x\}, \text{ there exist } x^0, x^1, \dots, x^m \in X, \text{ with } m \in \mathbb{Z}_{++}, \text{ such that} \\ \left(x^k, x^{k+1}\right) \in T_R \text{ for } k = 0, \dots, m-1, \text{ with } x^0 = x \;\&\; x^m = y. \end{array}$$

An *SCC* $F^{TC}$ on $\mathcal{P}^n$ is the *top-cycle SCC* if for all $R \in \mathcal{P}^n$, $F^{TC}(R) = TC(R)$.

The next proposition shows that $F^{TC}$ is partially-honest implementable, while it cannot be partially-honest implemented by any $s$-mechanism.

**Proposition 3.** *Let* Assumption 1 *hold* and $\mathcal{H} = 2^N\setminus\varnothing$. *(i)* $F^{TC}$ *is* partially-honest implementable; *(ii)* $F^{TC}$ *is not* partially-honest implementable by any $s$-mechanism.

**Proof.** Observe that Condition $\mu^{**}$(i) is vacuously satisfied by any *SCC*. Then, to see that $F^{TC}$ is partially-honest implementable, it suffices to observe that $F^{TC}$ satisfies the requirement of no-veto power which, in turn, implies Conditions $\mu^{**}$(ii)-$\mu^{**}$(iv). This completes part (i) of the statement.

To show part (ii), assume, to the contrary, that $F^{TC}$ is partially-honest implementable by an $s$-mechanism. Then, $F^{TC}$ satisfies Condition $M^*_s$, and, in particular, Condition $M^*_s$(i).

[24] $\mathcal{P}^n \subseteq \mathcal{R}^n$ is the set of all available profiles of linear orders.

[25] A relation $T$ on $X$ is a tournament if it is complete and asymmetric.

Let $N = \{1, 2, 3\}$ with $\#N = 3$, $X = \{x, y, z\}$, with $\#X = 3$, and $\mathcal{R}^3 = \{R, R^*\}$, where agents' preferences are as follows:

| $R$ | | | $R^*$ | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 1 | 2 | 3 |
| $x$ | $y$ | $z$ | $x$ | $y$ | $x$ |
| $y$ | $z$ | $x$ | $y$ | $z$ | $z$ |
| $z$ | $x$ | $y$ | $z$ | $x$ | $y$ |

With abuse of notation, we write $xT_Ry$ for $(x, y) \in T_R$. In terms of the tournament relation, we have that $xT_RyT_RzT_Rx$, while $xT_{R^*}a$ for all $a \in \{y, z\}$ and $yT_{R^*}z$. Since $y \in TC(R) = X$, there exists a profile of sets $(C_\ell(R_\ell, y))_{\ell \in N}$ such that $y \in C_\ell(R_\ell, y) \subseteq L(R_\ell, y) \cap X$ for all $\ell \in N$. Since $(R_\ell, R_{\ell+1}) \neq (R^*_\ell, R^*_{\ell+1})$ for $\ell \in \{2, 3\}$, it follows that Condition $M^*_s$(i) is vacuously satisfied if $H \cap \{2, 3\} \neq \varnothing$. The only case that we are left to verify is $H = \{1\}$. Since $(R_1, R_2) = (R^*_1, R^*_2)$ and $L(R_\ell, y) = L(R^*_\ell, y)$ for all $\ell \in N$, Condition $M^*_s$(i) implies that $y \in F^{TC}(R^*) \neq TC(R^*) = \{x\}$, a contradiction. ■

Before closing this section, it is important to note that the *Walrasian correspondence* and the *egalitarian-equivalent solution* (Pazner and Schmeidler, 1978), defined in the classical exchange economies, are other well-known examples of non-monotonic *SCC*s. These *SCC*s are not partially-honest implementable by any $s$-mechanism either, though they are partially-honest implementable by virtue of Theorem 2.

# 5 Two-agent implementation problems

Seminal papers on two-agent implementation are those of Moore and Repullo (1990) and Dutta and Sen (1991), who independently refined Maskin's characterization result (Maskin, 1999) by providing necessary and sufficient conditions for an *SCC* to be implementable.[26] Since Dutta and Sen's Condition $\beta$ and Moore and Repullo's Condition $\mu 2$ coincide in substance, we state only Condition $\mu 2$.

Condition $\mu 2$ (for short, $\mu 2$): There exists a set $Y \subseteq X$ and, for all $R \in \mathcal{R}^n$ and all $x \in F(R)$, there exists a profile of sets $(C_\ell(R, x))_{\ell \in N}$ such that $x \in C_\ell(R, x) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$; furthermore, Condition $\mu$ holds; finally, for all $R^* \in \mathcal{R}^n$, the following (iv) is satisfied:
(iv) for each $(x', R') \in X \times \mathcal{R}^2$ with $x' \in F(R')$,
(a) there exists an $e \equiv e(x', R', x, R) \in C_1(R', x') \cap C_2(R, x)$, with $e(x, R, x, R) = x$;
(b) if $C_1(R', x') \subseteq L(R^*_1, e)$ and $C_2(R, x) \subseteq L(R^*_2, e)$, then $e \in F(R^*)$.

Condition $\mu 2$ is markedly stronger than Condition $\mu$, as it includes a punishment condition - Condition $\mu 2$(iv). While the first part of Condition $\mu 2$(iv) guarantees the existence of a punishment outcome, the second part requires that if the punishment outcome is an equilibrium outcome, it should be $F$-optimal.

In the next two sub-sections, we identify the class of partially-honest implementable *SCC*s, not only in the case where the planner knows that exactly one agent is partially-honest, but also in the case where the exact number of partially-honest agents is unknown to her - Assumption 1. We present two new conditions which are not only necessary and sufficient conditions for *SCC*s to be partially-honest implementable, but also markedly weaker

[26] See also Busetto and Codognato (2009).

than Condition $\mu 2$. Significantly - and in line with earlier results and Theorem 2 -, our characterizations confirm that when agents hold preferences for truth-telling, the scope of implementation is enlarged. Yet, limits still remain. Particularly, what still limits implementability are the weaker variants of Condition $\mu 2$(v) embedded in our conditions on implementation. Sub-section 5.3 reports briefly the implications of our results.

## 5.1 Exactly one partially-honest agent

In this sub-section, we make the informational assumption that there exists exactly one partially-honest agent in society. The planner knows that there exists a dishonest averse agent but not who she is.

For the same reason highlighted in sub-section 3.1, Condition $\mu 2$ is not a necessary condition for partially-honest implementation. We amend this condition in the following way.

CONDITION $\mu 2^*$ (for short, $\mu 2^*$): Conditions $\mu^*$ holds; moreover, for all $H \in \mathcal{H}$, and for all $R^* \in \mathcal{R}^2$, the following condition (iv) is satisfied:
(iv) for each $(x', R') \in X \times \mathcal{R}^2$ with $x' \in F(R')$,
(a) there exists an $e \equiv e(x', R', x, R) \in C_1(R', x') \cap C_2(R, x)$, with $e(x, R, x, R) = x$;
(b) if $x' \neq x$, $R' \neq R$, $C_1(R', x') \subseteq L(R_1^*, e)$, $C_2(R, x) \subseteq L(R_2^*, e)$, and
(b.1) if $H = \{1\}$ and $\{e\} = \max_{R_1^*} C_1(R', x')$, then $e \in F(R^*)$;
(b.2) if $H = \{2\}$ and $\{e\} = \max_{R_2^*} C_2(R, x)$, then $e \in F(R^*)$.

In the next theorem, we show that the above Condition $\mu 2^*$ is necessary for implementation when exactly one agent holds preferences for truth-telling.

**Theorem 7.** *Let Assumption 1 hold and $\mathcal{H} = \{\{1\}, \{2\}\}$. If an SCC $F \in \mathcal{F}$* defined on $\mathcal{R}^2$ *is* partially-honest implementable, *then it satisfies* Condition $\mu 2^*$.

**Proof.** Let Assumption 1 hold and let $\mathcal{H} = \{\{1\}, \{2\}\}$. Let $\gamma \equiv (M, g)$ be a mechanism which partially-honest implements $F \in \mathcal{F}$, which is defined on $\mathcal{R}^2$. The proof that $F$ satisfies Condition $\mu^*$ follows from Theorem 1. Finally, we show that $F$ meets Condition $\mu 2^*$(iv). Take any $H \in \mathcal{H}$. Take any $(x', R', x, R) \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$ with $x \in F(R)$ and $x' \in F(R')$. Then, there exists an equilibrium strategy $m \equiv (m_1, m_2) \in NE\left(\gamma, \succcurlyeq^R\right)$ such that $g(m) = x$. Similarly, $m' \equiv (m_1', m_2') \in NE\left(\gamma, \succcurlyeq^{R'}\right)$ and $g(m') = x'$. Let $e \equiv e(x', R', x, R) = g(m_1, m_2')$. Then, defining $C_1(R', x') \equiv g(M_1, m_2')$ and $C_2(R, x) \equiv g(m_1, M_2)$, $e \in C_1(R', x') \cap C_2(R, x)$ holds, as sought. Finally, it is also clear that $F$ satisfies Condition $\mu 2^*$(iv.b) as, for instance, in the case of $\mu 2^*$(iv.b.1), if $e \notin F(R^*)$, then the only deviator is the partially-honest agent 1, but her deviation to an $m_1^* \in T_1^\gamma(R^*, F)$ results in the same outcome $e$ because $\{e\} = \max_{R_1^*} C_1(R', x')$, which is a contradiction. Thus, $F$ satisfies $\mu 2^*$(iv). ■

Though Condition $\mu 2^*$ is a necessary condition for partially-honest implementation, it is too weak to guarantee the sufficiency result. To this end, other requirements exist. These requirements are that the domain of preferences must be large enough, and that $F$ satisfies Condition $\mu^{**}$ and an extra auxiliary condition. The condition as a whole can be stated as follows.

CONDITION $\mu 2^{**}$ (for short, $\mu 2^{**}$): Condition $\mu^{**}$ holds;[27] moreover, for all $H \in \mathcal{H}$, and for

[27] We refer to the condition that requires only one of the statements (i)–(iv) in Condition $\mu^{**}$ as Conditions $\mu 2^{**}$(i)–$\mu 2^{**}$(iv) each.

all $R^* \in \mathcal{R}^2$, the following condition (v) is satisfied:
(v) for each $(x', R') \in X \times \mathcal{R}^2$ with $x' \in F(R')$,
(a) there exists an $e \equiv e(x', R', x, R) \in C_1(R', x') \cap C_2(R, x)$, with $e(x, R, x, R) = x$;
(b) if $x' \neq x$, $R' \neq R$, $C_1(R', x') \subseteq L(R_1^*, e)$, $C_2(R, x) \subseteq L(R_2^*, e)$, and $e \notin F(R^*)$, then;
(b.1) if $R = R^*$, then $H = \{2\}$;
(b.2) if $R' = R^*$, then $H = \{1\}$;
(c) if $R = R' = R^*$, $x' \neq x$, $(e, x') \in I_1^*$, and $(e, x) \in I_2^*$, then $e \in F(R^*)$.

The next theorem shows that this condition is not only sufficient, but also necessary for partially-honest implementation, when only game forms with simple punishment are admissible (the formal proof is relegated to Appendix).

**Theorem 8.** *Let Assumption 1,* $\Gamma = \Gamma_{SP}$*, and* **RD** *hold, and let* $\mathcal{H} = \{\{1\}, \{2\}\}$*. An SCC* $F \in \mathcal{F}$ defined on $\mathcal{R}^2$ *is* partially-honest implementable *if and only if it satisfies* Condition $\mu 2^{**}$.

## 5.2 There are partially-honest agents

This sub-section makes the informational assumption that the planner knows that there are partially-honest agents, but she knows neither their identities nor their exact number. Its objective is to fully identify the class of partially-honest implementable *SCC*s under this informational assumption.

To this end, as done in the previous sub-section, let us lay down the condition that every *SCC F* must meet if it is partially-honest implementable. The condition can be stated as follows.

Condition $\mu 2^\circ$ (for short, $\mu 2^\circ$): Condition $\mu 2^*$ holds; moreover, for all $R^* \in \mathcal{R}^2$, the following condition (v) is satisfied:
(v) for all $i \in N$ and all $H \in \mathcal{H}$, if $H = N$, $R = R^*$, $y \in C_i(R, x) \subseteq L(R_i^*, y)$, and $y \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, then $y \in F(R^*)$ whenever $x = y$.

It is easy to confirm that Condition $\mu 2^\circ$(v) is necessary. By virtue of Theorem 7, the next theorem states that Condition $\mu 2^\circ$ is necessary for partially-honest implementation, while omitting the proof of it.

**Theorem 9.** *Let Assumption 1. If an SCC* $F \in \mathcal{F}$ defined on $\mathcal{R}^2$ *is* partially-honest implementable, *then it satisfies* Condition $\mu 2^\circ$.

Condition $\mu 2^\circ$ alone does not suffice to guarantee partial-honest implementation. Let us strengthen it as follows.

Condition $\mu 2^{\circ\circ}$ (for short, $\mu 2^{\circ\circ}$): Condition $\mu 2^{**}$ holds;[28] moreover, for all $R, R^* \in \mathcal{R}^2$, the following condition (vi) is satisfied:
(vi) for all $H \in \mathcal{H}$, all $x \in F(R)$, and all $i \in N$, if $H = N$, $R = R^*$, $y \in C_i(R, x) \subseteq L(R_i^*, y)$, and $y \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, then $y \in F(R^*)$.

This condition guarantees the sufficiency result when the domain of preferences is sufficiently rich. However, to close the gap between what constitutes a necessary and a sufficient condition, we focus on game forms which satisfy the following stronger variant of punishment condition.

[28] We refer to the condition that requires only one of the statements (i)–(v) in Condition $\mu 2^{**}$ as Conditions $\mu 2^{\circ\circ}$(i)–$\mu 2^{\circ\circ}$(v) each.

**Strong Punishment (StP):** For any $R, R' \in \mathcal{R}^2$, any $i \in N$, and any $m \equiv (m_i, m_\ell) \in M$ such that $g(m) = x$, there exists an $m'_i \in T_i^\gamma(R', F)$ such that $g(m'_i, m_\ell) = g(m)$.

A mechanism $\gamma$ is a *mechanism with strong punishment* if it satisfies **StP**. Denote the class of mechanisms satisfying **StP** by $\Gamma_{StP}$.

The above condition has a similar flavor to **SP**. However, with condition **StP**, the planner is required to design a game form in which if $x$ is an attainable outcome at state $R$ - in the sense that there is a message profile $m$ leading to it under this state -, then an agent $i$ should be able to reach this $x$ by replacing the untruthful message $m_i$ with a truthful one $m'_i$ (while keeping constant the messages of all others). Therefore, differently from **SP**, every attainable outcome can be supported by a truthful message profile, regardless of whether it is an $F$-optimal outcome. In this sense, the above condition can be considered a strong punishment requirement. Similar to **SP**, the requirement of **StP** is satisfied by all classical mechanisms in the literature of Nash implementation (see, for instance, Repullo, 1987; Moore and Repullo, 1990; Saijo, 1988; Dutta and Sen, 1991; Tatamitani, 2001).

The following theorem shows that Condition $\mu 2^{\circ\circ}$ is necessary and sufficient for partially-honest implementation, when the domain of preferences is sufficiently rich and the focus is on mechanisms with strong punishment (the formal proof is relegated to Appendix).

**Theorem 10.** *Let Assumption 1,* $\Gamma = \Gamma_{StP}$, *and* **RD** *hold. An SCC* $F \in \mathcal{F}$ defined on $\mathcal{R}^2$ *is* partially-honest implementable *if and only if it satisfies* Condition $\mu 2^{\circ\circ}$.

Before closing this sub-section, it may be worth mentioning briefly that if the planner knows that both agents are partially-honest, the class of partially-honest implementable *SCC*s becomes larger, since neither Condition $\mu 2^{**}$(ii), Condition $\mu 2^{**}$(iv), nor Condition $\mu 2^{**}$(v.b) is required. This result is readily obtained by Theorem 10.

**Corollary 3.** *Let Assumption 1 and* $\mathcal{H} = \{N\}$. *An SCC* $F \in \mathcal{F}$ defined on $\mathcal{R}^2$ *is* partially-honest implementable by a mechanism in $\Gamma_{StP}$ *if and only if it satisfies* Condition $\mu 2^{\circ\circ}$ *without* Condition $\mu 2^{**}$(ii), Condition $\mu 2^{**}$(iv), *or* Condition $\mu 2^{**}$(v.b)

Notice that the above result does not postulate any requirement on the domain of preferences.

## 5.3 Implications

Condition $\mu 2^{\circ\circ}$ - and so Condition $\mu 2^{**}$ - imposes non-trivial restrictions on $F$. For example, the *Pareto SCC* is not partially-honest implementable by virtue of Proposition 1. Despite this, the results of the above sub-sections are quite permissive.[29] In the following, we justify this assertion by considering economically meaningful assumptions on the domain of preferences and on the set of outcomes.

There are economic environments in which the following assumption is satisfied.

**Assumption 2** (Moore and Repullo, 1990, p. 1093). *There exists a* bad outcome $b \in X$ *such that for all* $R \in \mathcal{R}^2$ *and* $i \in N$, $(x, b) \in P_i$ *for all* $x \in F(\mathcal{R}^2) \equiv \{y \in X | y \in F(R')$ for some $R' \in \mathcal{R}^2\}$.

[29] For a non-dictatorial and weakly Pareto efficient partially-honestly implementable *SCC* defined on the domain of linear orders which rebuts the negative conclusion of Hurwicz and Schmeidler (1978), we refer the reader to Dutta and Sen (2009).

For example, consider an exchange economy in which agents have strongly monotonic preferences and the *SCC* assigns only positive consumption bundles. Under free disposal, one can define the null consumption bundle as the bad outcome.

If there is a bad outcome, we can set $e(x, R, x', R') = b$ for each $(x, R, x', R') \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$ to satisfy Condition $\mu 2^{\circ\circ}$(v) vacuously. Moreover, though Condition $\mu 2^{\circ\circ}$ can be checked by using the algorithm provided by Sjöström (1991), a condition which can be promptly checked is stated below.

RESTRICTED VETO POWER (Moore and Repullo, 1990, p. 1093): For all $i \in N$, all $R \in \mathcal{R}^2$, all $x \in X$, and all $x' \in F(\mathcal{R}^2) \equiv \{y \in X | y \in F(R) \text{ for some } R \in \mathcal{R}^2\}$, if $x \in \max_{R_\ell} X$ for all $\ell \in N \setminus \{i\}$ and $(x, x') \in R_i$, then $x \in F(R)$ holds.

The next result shows that the condition of restricted veto power suffices to guarantee partially-honest implementation when Assumptions 1 and 2 hold.

**Corollary 4.** *Let Assumption 1 and Assumption 2 hold. An SCC $F$* on $\mathcal{R}^2$ *is* partially-honest implementable *if it satisfies* restricted veto power.

**Proof.** Let Assumption 1 and Assumption 2 hold. Suppose that $F$ on $\mathcal{R}^2$ satisfies restricted veto power. It suffices to show that Assumption 2 and restricted veto power imply Condition $\mu 2^{\circ\circ}$. Let $Y = X$, and for all $R \in \mathcal{R}^2$ and all $x \in F(R)$, let $C_i(R, x) = L(R_i, x)$ for all $i \in N$. Since Assumption 2 holds, for each $(x, R, x', R') \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$ with $x \in F(R)$ and $x' \in F(R')$, let $e(x', R', x, R) = b$ if $(x, R) \neq (x', R')$; otherwise, $e(x, R, x', R') = x$. Then, Condition $\mu 2^{\circ\circ}$(v) is satisfied. Since restricted veto power implies Conditions $\mu$(ii)-$\mu$(iii) which, in turn, imply Conditions $\mu 2^{\circ\circ}$(ii)-$\mu 2^{\circ\circ}$(iv) and Condition $\mu 2^{\circ\circ}$(vi), the statement follows. ■

By Corollary 3, an *SCC* is partially-honest implementable by Condition $\mu 2^{\circ\circ}$ without Conditions $\mu 2^{\circ\circ}$(ii), $\mu 2^{\circ\circ}$(iv), or $\mu 2^{\circ\circ}$(v.b) if the planner knows that both agents are partially-honest. Under this informational assumption, unanimity and a weakening of restricted veto power, when combined with Assumption 2, suffice to guarantee partially-honest implementation.[30] The weakening of restricted veto power can be stated as follows.

WEAK RESTRICTED VETO POWER: For all $i \in N$, all $R \in \mathcal{R}^2$, all $x \in X$, and all $x' \in F(R)$, if $x \in \max_{R_\ell} X$ for all $\ell \in N \setminus \{i\}$ and $(x, x') \in R_i$, then $x \in F(R)$ holds.

We can now state the following result.

**Corollary 5.** *Let Assumption 1 and Assumption 2 hold, and let $\mathcal{H} = \{N\}$. An SCC $F$* on $\mathcal{R}^2$ *is* partially-honest implementable *if it satisfies* weak restricted veto power *and* unanimity.

**Proof.** It is obvious, so omitted. ■

Suppose that two agents bargain over the division of one unit of a perfectly divisible good. If they do not reach an agreement, they both receive nothing. In this framework, non-monotonic strongly individual-rational bargaining solutions[31] defined on the class of utility possibility sets - such as the *Nash bargaining solution* - are special examples of *SCC*s applied to Corollary 4 and Corollary 5, setting the disagreement point $d = (0, 0)$ as a bad outcome.[32]

[30] For the definition of unanimous SCCs, see section 2.

[31] A bargaining solution is strongly individual-rational if it provides agents with agreements which give them utilities higher than those they derive from the disagreement point $d$.

[32] For the Nash bargaining solution defined on the class of utility possibility sets, see Vartiainen (2007).

Finally, let us consider an interesting domain restriction. Before defining it, let $SL(R_i, x)$ denote agent $i$'s strict lower contour set at $(R_i, x) \in \mathcal{R}_i \times X$; that is, $SL(R_i, x) \equiv \{y \in X | (x, y) \in P_i\}$. The assumption on the domain of preferences can be defined as follows.

**Assumption 3** (Busetto and Codognato, 2009). *$\mathcal{R}^2$ is such that for all $R^* \in \mathcal{R}^2$, we have*:
(i) $\max_{R_i^*} SL(R_i, x) \cap \max_{R_j^*} SL(R_i, x) = \varnothing$ for all $i, j \in N$, with $i \neq j$, all $R \in \mathcal{R}^2$, and all $x \in X$;
(ii) $\max_{R_1^*} SL(R_1', x') \cap \max_{R_2^*} SL(R_2, x) = \varnothing$ for each $(x, R, x', R') \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$, with $(x, R) \neq (x', R')$.

This domain restriction is very mild and much weaker than *Assumption E* imposed by Moore and Repullo (1990, p. 1095) and *Assumptions 5.1-5.2* imposed by Dutta and Sen (1991, p. 125), whenever $X$ is a subset of a finite-dimensional Euclidean space.[33] For example, this restriction is satisfied in environments with continuous and locally non-satiated preferences or in environments in which the set of outcomes is a space of lotteries over a finite set of outcomes and agents' preferences over lotteries are represented by von Neumann-Morgenstern utility functions. Given Assumption 3, we can define a condition that, when combined with others, suffices to ensure Condition $\mu 2^{\circ\circ}$.

**Definition 7.** An *SCC* $F$ on $\mathcal{R}^2$ satisfies the *non-empty lower intersection* if for all $(x, R, x', R') \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$, with $x \in F(R)$ and $x' \in F(R')$, we have that $SL(R_1', x') \cap SL(R_2, x) \neq \varnothing$.

This property appears in Moore and Repullo (1990) and Dutta and Sen (1991) and holds in many environments. For example, it holds in an exchange economy for which indifference curves never touch the axes, and for which the *SCC* recommends only interior allocations. We can now state our last two results.

**Corollary 6.** *Let Assumption 1 and Assumption 3 hold. An SCC $F$ on $\mathcal{R}^2$ is* partially-honest implementable *if it satisfies* non-empty lower intersection *and* restricted veto power.

**Proof.** Let Assumption 1 and Assumption 3 hold. Suppose that $F$ on $\mathcal{R}^2$ satisfies non-empty lower intersection, weak restricted veto power, and unanimity. We show that $F$ is partially-honest implementable. It suffices to show that Condition $\mu 2^{\circ\circ}$ is implied by our suppositions.

For all $i \in N$, all $(x, R) \in X \times \mathcal{R}^2$, and all $x \in F(R)$, let $C_i(R, x) = SL(R_i, x) \cup \{x\}$ and $Y = X$. It is easy to verify that $C_i(R, x) \subseteq L(R_i, x) \cap Y$. For all $(x', R', x, R) \in X \times \mathcal{R}^2 \times X \times \mathcal{R}^2$, with $x \in F(R)$ and $x' \in F(R')$, let $e(x', R', x, R) \in SL(R_i', x') \cap SL(R_\ell, x)$ if $(x, R) \neq (x', R')$; otherwise, $e(x', R', x, R) = x$. By definition of $e(x', R', x, R)$ and non-empty lower intersection, it is easy to see that Condition $\mu 2^{\circ\circ}$(v.a) is satisfied, while Condition $\mu 2^{\circ\circ}$(v.b) and Condition $\mu 2^{\circ\circ}$(v.c) are vacuously satisfied since $(x, R) \neq (x', R')$.

Take any $R, R^* \in \mathcal{R}^2$. Let $x \in F(R)$, $y \in C_i(R, x) \subseteq L(R_i^*, y)$, and $y \in \max_{R_\ell^*} Y$ for $i, \ell \in N$ with $i \neq \ell$. It cannot be that $y \in C_i(R, x) \setminus \{x\}$; otherwise, $y \in \max_{R_i^*} SL(R_i, x) \cap \max_{R_j^*} SL(R_i, x)$, contradicting Assumption 3(i). Let $x = y$. Condition $\mu 2^{\circ\circ}$(vi) is satisfied, trivially. Moreover, since restricted veto power implies that $x \in F(R^*)$, it follows that Condition $\mu 2^{\circ\circ}$(ii) and Condition $\mu 2^{\circ\circ}$(iv) are satisfied. Clearly, restricted veto power implies Condition $\mu 2^{\circ\circ}$(iii). The statement follows by observing that Condition $\mu 2^{\circ\circ}$(i) is satisfied by all $F \in \mathcal{F}$. ■

[33] The formal arguments are provided in Busetto and Codognato (2009).

**Corollary 7.** *Let Assumption 1 and Assumption 3 hold, and let* $\mathcal{H} = \{N\}$*. An SCC* $F$ on $\mathcal{R}^2$ *is* partially-honest implementable *if it satisfies* non-empty lower intersection *and* unanimity.

**Proof.** The proof of this statement directly follows from the proof of Corollary 6, and so it is omitted here. ■

Consider a two-agent exchange economy with $\ell \geq 2$ divisible goods, in which agents have continuous and strongly monotonic preferences, and in which indifference curves never touch the axes (for instance, Cobb-Douglas preferences). Suppose that an *SCC* $F$ selects only interior allocations of the feasible set. In this setting, restricted veto power, unanimity, and non-empty lower intersection are satisfied by this $F$. The *Walrasian correspondence* and the *egalitarian-equivalent solution* are examples of such *SCC*s.[34] This implies that they are partially-honest implementable, according to Corollary 6 and Corollary 7.

# 6 Concluding remarks

In this closing section, rather than restating the main contributions of the paper, we conclude with a word of caution and with a couple of alleys for research.

In the framework developed by Moore and Repullo (1990), this paper has studied the consequences of injecting a minimal dishonesty aversion in implementation theory. While it is undeniable that there are people who care not only about welfaristic features of the consequences, but also - to some extent - non-consequential features of lying, it is equally undeniable that it would be a mistake to apply the kind of aversion studied here carelessly. Caution seems advisable in all applied fields in which the idea of partial-honesty may not be appealing or plausible, like in the playground of auction design. Nonetheless, this idea can be fruitfully applied to a wide range of public decision making problems. Applications to problems of public goods provision, externalities, voting, taxation, and income distribution seem to hold exciting potential. The tools developed and the results reported herein can provide useful arguments and insights in this respect.

Second, while the paper has focussed on a minimal aversion to lying by agents involved in a mechanism, the departure from the standard assumption that agents are unconcerned about the non-welfaristic features of the consequences can be modelled in a variety of ways. An interesting direction has been taken up in a recent work by Lombardi and Yoshihara (2011b), where the authors explore the consequences of injecting a 'stronger' degree of honesty in implementation problems by also connecting the outcome announcement with the deception. It is certainly worth considering other views on modelling agents' preferences.

Third, while a considerable amount of experimental data suggests that agents may display preferences for truth-telling, all lab experiments designed to test whether or not agents consider more than "just" their material payoffs in strategic situations are not geared towards implementation theory. There is little evidence that experimental subjects are willing to uphold the truth when called to perform implementation tasks if consequences of doing so are not costly - e.g., Cabrales *et al.* (2003). The design of experimental tests for dishonesty aversion specifically tailored towards implementation theory is highly desirable and promise to be a fruitful and interesting area of research for years to come.

[34] This non-monotonic SCC is well-defined under our assumptions on preferences (Pazner and Schmeidler, 1978).

Finally, while the paper sets solid foundations for implementation with partially-honest agents, it falls short in many important aspects. For example, while the paper specified the set of properties that an *SCC* should satisfy in order to be partially-honest implementable, the devised mechanisms present the disadvantage of involving complex strategy spaces. In particular, strategies include either whole preference profiles or whole indifference sets of several agents. This implies that the message space is of infinite dimension in many economic applications. Furthermore, the components of the strategy space do not have a straightforward economic interpretation such as consumption bundles, allocations, and prices. Therefore, there is a need for specifying the scope of the analysis reported herein away from abstract social choice environments. In this regard, the exploration of the rich set of implications that arise from the injection of a minimal dishonesty aversion to economic agents involved in a mechanism can take many directions. One interesting direction is explored in a recent work of Lombardi and Yoshihara (2011a) in which implementation of efficient *SCC*s by natural mechanisms is analyzed in classical exchange economies and results in line with those reported herein are unveiled.

# References


[1] Bergemann, D., Morris, S., and O. Tercieux (2010): Rationalizable implementation. Accepted for publication at the *Journal of Economic Theory*.

[2] Busetto, F. and G. Codognato (2009): Reconsidering two-agent Nash implementation. *Social Choice and Welfare*, 32, 171-179.

[3] Cabrales, A., G. Charness and L.C. Corchon (2003): An experiment on Nash implementation. *Journal of Economic Behavior and Organization*, 51, 161-193.

[4] Cabrales, A. and R. Serrano (2010): Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms. Accepted for publication at *Games and Economic Behavior*.

[5] Camerer, C.F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, New Jersey, USA.

[6] Danilov, V. (1992): Implementation via Nash equilibria. *Econometrica*, 60, 43-56.

[7] Dutta, B. and A. Sen (1991): A necessary and sufficient condition for two-person Nash implementation. *Review of Economic Studies*, 58, 121-128.

[8] Dutta, B. and A. Sen (2009): Nash implementation with partially-honest individuals. *Warwick Economic Research Papers*, n. 920, Warwick University. Accepted for publication at *Games and Economic Behavior*.

[9] Dutta, B., Sen, A., and Vohra, R. (1995): Nash implementation through elementary mechanisms in economic environments. *Economic Design*, 1, 173-204.

[10] Eliaz, K. (2002): Fault-Tolerant Implementation. *Review of Economic Studies*, 69, 589-610.

[11] Glazer, J., and A. Rubinstein (1998): Motives and Implementation: On the Design of Mechanisms to Elicit Opinions. *Journal of Economic Theory*, 79, 157-173.

[12] Gneezy, U. (2005): Deception: The Role of Consequences. *American Economic Review*, 95, 384–394.

[13] Hurkens, S. and N. Kartik (2009): Would I Lie to You? On Social Preferences and Lying Aversion. *Experimental Economics*, **12**, 180-192.

[14] Hurwicz, L. (1960): Optimality and informational efficiency in resource allocation processes. In: Arrow, K.J., Karlin, S., Suppes, P. (eds), *Mathematical Methods in the Social Sciences*. Stanford University Press, 27-46.

[15] Hurwicz, L. (1972): On Informationally Decentralized Systems," in Decision and Organization, ed. by R. Radner and C. B.McGuire, North-Holland, 297–336.

[16] Hurwicz, L. (1986): On the Implementation of Social Choice Rules in Irrational Societies. In *Social Choice and Public Decision Making*, Essays in Honor of Kenneth J. Arrow Volume I, ed. by W. Heller, R. Starr and D. Starrett. Cambridge University Press, USA.

[17] Hurwicz, L. and D. Schmeidler (1978): Outcome functions which guarantee the existence and Pareto optimality of Nash equilibria. *Econometrica*, 46, 144-174.

[18] Jackson, M. (2001): A crash course in implementation theory. *Social Choice and Welfare*, 18, 655-708.

[19] Lombardi, M. and N. Yoshihara (2010): A full characterization of Nash implementation with strategy space reduction. Available at SSRN: http://ssrn.com/abstract=1593690.

[20] Lombardi, M. and N. Yoshihara (2011a): Natural implementation with partially-honest agents. *Mimeo in progress*, Hitotsubashi University.

[21] Lombardi, M. and N. Yoshihara (2011b): On partially-honest Nash implementation by self-relevant mechanisms: A stronger definition of partial-honesty. *Mimeo in progress*, Hitotsubashi University.

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

[23] Maskin, E. and T. Sjöström (2002): The theory of implementation. In *Handbook of Social Choice and Welfare*, Vol. 1, K. Arrow, A.K. Sen and K. Suzumura, eds. Amsterdam: Elsevier Science.

[24] Matsushima, H. (2008): Role of honesty in full implementation. *Journal of Economic Theory*, 139, 353-359.

[25] Moore, J., and R. Repullo (1990): Nash implementation: A full characterization. *Econometrica*, 58, 1083-1100.

[26] Pazner, E. and D. Schmeidler (1978): Egalitarian equivalent allocations: A new concept of economic equity. *Quarterly Journal of Economics*, 92, 671-687.

[27] Renou, L. and K.H. Schlag (2011): Minimax regret implementation. *Games and Economic Behavior*, 71, 527-533.

[28] Repullo, R. (1987): A Simple Proof of Maskin Theorem on Nash Implementation. *Social Choice and Welfare*, 4, 39-41.

[29] Sen, A.K. (1997): Maximization and the act of choice. *Econometrica*, 65, 745-779.

[30] Saijo, T. (1988): Strategy space reduction in Maskin's theorem: sufficient conditions for Nash implementation. *Econometrica*, 56, 693-700.

[31] Saijo, T., Tatamitani, Y., and Yamato, T. (1996): Toward natural implementation, *International Economic Review*, 37, 949-980.

[32] Sjöström, T. (1991): On the necessary and sufficient conditions for Nash implementation. *Social Choice and Welfare*, 8, 333-340.

[33] Tatamitani, Y. (2001): Implementation by self-relevant mechanisms. *Journal of Mathematical Economics*, 35, 427-444.

[34] Tatamitani, Y. (2002): Implementation by self-relevant mechanisms: application. *Mathematical Social Science*, 44, 253-276.

[35] Vartiainen, H. (2007): Nash implementation and bargaining problem. *Social Choice and Welfare*, 29, 333-351.

[36] Yamato, T. (1992): On Nash implementation of social choice correspondences. *Games and Economic Behavior*, 4, 484-492.

# 7 Appendix

**Proof of Theorem 2.** Let Assumption 1 hold and let $\mathcal{R}^n$ satisfy **RD**. Take any $F \in \mathcal{F}$. Let $\diamond \in N$ be an arbitrary agent index.

*1. The necessity of Condition* $\mu^{**}$.

Let $F$ be partially-honest implemented by $\gamma \equiv (M, g) \in \Gamma_{SP}$. Let $Y \equiv g(M)$. Take any $R \in \mathcal{R}^n$ and any $x \in F(R)$. Then, there exists an $m(R, x) \in NE\left(\gamma, \succcurlyeq^R\right)$ for any $H' \in \mathcal{H}$, such that $g(m(R, x)) = x$ and $m_h(R, x) \in T_h^\gamma(R, F)$ for any $h \in H'$ and any $H' \in \mathcal{H}$, because $\gamma \in \Gamma_{SP}$. For all $\ell \in N$, let $C_\ell(R, x) \equiv g(M_\ell, m_{-\ell}(R, x))$. Then, $C_\ell(R, x) \equiv g(M_\ell, m_{-\ell}(R, x)) \subseteq L(R_\ell, x) \cap Y$ for all $\ell \in N$. Take any $R^* \in \mathcal{R}^n$ and any $H \in \mathcal{H}$. By Theorem 1, it follows that $F$ satisfies Condition $\mu^*$. Thus, we only show that $F$ satisfies $\mu^{**}$(ii.c)-$\mu^{**}$(iv).

Given $i \in N$, suppose that $y \in C_i(R, x) \subseteq L(R_i^*, y)$, $y \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, and $y \notin F(R^*)$. Thus, $g(m_i, m_{-i}(R, x)) = y$ for some $m_i \in M_i$. Assume, to the contrary, that $R = R^*$ and $i \notin H$. Then, $(m_i, m_{-i}(R, x)) \in NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for this specific $H$, a contradiction. Hence, $F$ satisfies Condition $\mu^{**}$(ii.c).

Suppose that $y \in \max_{R_\ell^*} Y$ for all $\ell \in N$. Then, there exists an $\bar{m} \in M$ such that $g(\bar{m}) = y$. Consider $\bar{R} \equiv \left(\bar{R}_\ell\right)_{\ell \in N} \in \mathcal{R}^n$ such that $L\left(\bar{R}_\ell, y\right) = L(R_\ell^*, y)$ with $\partial L\left(\bar{R}_\ell, y\right) = \{y\}$ for all $\ell \in N$. Since $\mathcal{R}^n$ satisfies **RD**, such a profile is admissible. Condition $\mu^*$(iii) implies that $y \in F\left(\bar{R}\right)$, given that $F$ satisfies Condition $\mu^*$. Suppose that there exists a non-empty set $S \subseteq N$ such that $\bar{m}_\ell \notin T_\ell^\gamma(R^*, F)$ for all $\ell \in S$; otherwise $g(\bar{m}) \in F(R^*)$, as sought. Then, by **SP**, for each $\ell \in S$, there exists an $\bar{m}'_\ell \in T_\ell^\gamma(R^*, F)$ such that $g(\bar{m}'_\ell, \bar{m}_{-\ell}) = y$. By repeatedly applying **SP** from $\ell_1 \in S$ to $\ell_s \in S$, where $S = \{\ell_1, \ldots, \ell_s\}$, it follows that $g(\bar{m}'_S, \bar{m}_{-S}) = y$. Thus, $(\bar{m}'_S, \bar{m}_{-S}) \in NE\left(\gamma, \succcurlyeq^{R^*}\right)$ for any $H' \in \mathcal{H}$. Therefore, $F$ satisfies Condition $\mu^{**}$(iii).

Take any $i \in N$. Suppose that $L(R_i, x) = L(R_i^*, x)$, $x \in \max_{R_\ell^*} Y$ for all $\ell \in N \setminus \{i\}$, $R_{-i} = R_{-i}^*$, and $x \notin F(R^*)$. Then, since $x = g(m(R, x))$ and $g(M_i, m_{-i}(R, x)) \subseteq L(R_i, x) = L(R_i^*, x)$, it follows from the implementability of $F$ that $R_i^* \neq R_i$ and $m(R, x) \notin NE\left(\gamma, \succcurlyeq^{R^*}\right)$ holds for any $H' \in \mathcal{H}$. It follows that there is an $h \in H$ such that $m_h(R, x) \notin T_h^\gamma(R^*, F)$ and $(g(m_h, m_{-h}(R, x)), g(m(R, x))) \in I_h^*$ for some $m_h \in T_h^\gamma(R^*, F)$. Assume, to the contrary, that $H = \{i\}$. Then, the only deviator is agent $i$. Since $\gamma$ satisfies **SP**, there exists an $m_i^* \in T_i^\gamma(R^*, F)$ such that $g(m_i^*, m_{-i}(R, x)) = g(m(R, x)) = x$. This implies that $(m_i^*, m_{-i}(R, x)) \in NE\left(\gamma, \succcurlyeq^{R^*}\right)$ and so $x \in NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for this $H = \{i\}$, a contradiction. Therefore, $F$ satisfies Condition $\mu^{**}$(iv).

*2. The sufficiency of Condition* $\mu^{**}$.

Suppose that $F$ satisfies Condition $\mu^{**}$. Let $\gamma \equiv (M, g)$ be the mechanism defined in sub-section 3.1. For each $\ell \in N$, the set of truthful message is that defined in (1). By construction, $\gamma \in \Gamma_{SP}$. Take any $R \in \mathcal{R}^n$.

To show that $F(R) \subseteq NA\left(\gamma, \succcurlyeq^R\right)$ for any $H \in \mathcal{H}$, let $x \in F(R)$ and suppose that, for all $\ell \in N$, $m_\ell = (R, x, \diamond) \in T_\ell^\gamma(R, F)$. *Rule 1* implies that $g(m) = x$. Suppose that $\ell \in N$ deviates from $m_\ell$ to $m_\ell^* \in M_\ell$. It follows from *Rules 2* that $g(M_\ell, m_{-\ell}) = C_\ell(R, x) \subseteq L(R_\ell, x)$. We conclude that $m \in NE\left(\gamma, \succcurlyeq^R\right)$ and so $x \in NA\left(\gamma, \succcurlyeq^R\right)$ for any $H \in \mathcal{H}$, since $m_\ell = (R, x, \diamond) \in T_\ell^\gamma(R, F)$ for each $\ell \in N$.

To show that $NA\left(\gamma, \succcurlyeq^R\right) \subseteq F(R)$ for any $H' \in \mathcal{H}$, taking any $H \in \mathcal{H}$, let $m \in NE\left(\gamma, \succcurlyeq^R\right)$ with $g(m) = x$ for this $H$, and let us consider the following cases.

*Case 1*: $m$ corresponds to *Rule 1*.

Suppose that $R \neq \bar{R} = R^{\ell}$ for all $\ell \in N \setminus \{i\}$. Then, $m_h \notin T_h^{\gamma}(R, F)$ for all $h \in H$. Take any $m'_h \in T_h^{\gamma}(R, F)$ such that the outcome announced is $x^h = x$. *Rule 2.2* implies that $g(m'_h, m_{-h}) = x$ so that $((m'_h, m_{-h}), m) \in \succ_h^R$, a contradiction. Otherwise, $R = \bar{R}$ and so $x \in F(R)$.

*Case 2*: $m$ corresponds to *Rule 2.1.*

Then, $Y \subseteq L(R_{\ell}, x)$ for all $\ell \in N \setminus \{i\}$ and $C_i(\bar{R}, x) \subseteq L(R_i, x)$. Suppose that $R^i = R^{\ell} = \bar{R} \neq R$. Let $i \notin H$ and there is another $h \in H$. Agent $h$ can induce *Rule 3* by unilaterally deviating to $m'_h = (R, x, k^h) \in T_h^{\gamma}(R, F)$. By choosing $k^h$ so as to have $h = \ell^*(m_{-h}, m'_h)$, she obtains $g(m_{-h}, m'_h) = x$. Then, $((m_{-h}, m'_h), m) \in \succ_h^R$, a contradiction. Otherwise, let $i \in H$. As agent $i$ can induce *Rule 2.2* by deviating to $m'_i = (R, x, \diamond) \in T_i^{\gamma}(R, F)$, we have that $g(m_{-h}, m'_h) = x$, which again leads to a contradiction. Therefore, $\bar{R} = R$ and so $x \in F(R)$.

*Case 3*: $m$ corresponds to *Rule 2.2.*

Then, $Y \subseteq L(R_{\ell}, g(m))$ for all $\ell \in N \setminus \{i\}$ and $C_i(\bar{R}, x) \subseteq L(R_i, g(m))$. Suppose that $m_h \notin T_h^{\gamma}(R, F)$ for some $h \in H \setminus \{i\}$. Then, agent $h \in H \setminus \{i\}$ can induce *Rule 3* by deviating to a suitable $m'_h \in T_h^{\gamma}(R, F)$ so as to obtain $g(m'_h, m_{-h}) = g(m)$, which leads to $m \notin NE(\gamma, \succcurlyeq^R)$, a contradiction. Therefore, $m_h \in T_h^{\gamma}(R, F)$ for all $h \in H \setminus \{i\}$. Suppose that $\#H > 1$ and $i \in H$. As $m_h \in T_h^{\gamma}(R, F)$ for all $h \in H \setminus \{i\}$, it follows that $R = \bar{R}$ and $x \in F(R)$. Since $m$ falls into *Rule 2.2*, it follows that $R^i \neq R$, so that $m_i \notin T_i^{\gamma}(R, F)$. It follows from $x \in C_i(R, x) \subseteq L(R_i, g(m))$ and $g(m) \in C_i(R, x) \subseteq L(R_i, x)$ that $(x, g(m)) \in I_i$. Agent $i$ can deviate to $m'_i = (R, x, k^i) \in T_i^{\gamma}(R, F)$ so that she induces *Rule 1* and obtains $g(m'_i, m_{-i}) = x$, which contradicts $m \in NE(\gamma, \succcurlyeq^R)$. Therefore, $\#H \not> 1$ or $i \notin H$. Suppose that $\#H \geq 1$ and $i \notin H$. Since $R = \bar{R}$, Condition $\mu^{**}$(ii.c) implies that $g(m) \in F(R)$. Otherwise, let $H = \{i\}$. Observe that $R \neq \bar{R} = R^{\ell}$ for all $\ell \in N \setminus \{i\}$; otherwise agent $i$ can induce *Rule 1* by deviating to a suitable truthful message and obtain a profitable deviation. Notice that $m_i \in T_i^{\gamma}(R, F)$, otherwise agent $i$ can induce *Rule 2.2* by deviating to an $m'_i = (R, g(m), k^i) \in T_i^{\gamma}(R, F)$ and obtain $g(m'_i, m_{-i}) = g(m)$, which leads to $m \notin NE(\gamma, \succcurlyeq^R)$, a contradiction. Take an $\hat{R}_i \in \mathcal{R}_i(X)$ such that $L\left(\hat{R}_i, g(m)\right) = L(R_i, g(m))$ with $\partial L\left(\hat{R}_i, g(m)\right) = \{g(m)\}$. As $\mathcal{R}^n$ satisfies **RD**, we have that $\hat{R} \equiv \left(\hat{R}_i, R_{-i}\right) \in \mathcal{R}^n$. Then, $\mu^{**}$(ii.a) implies that $g(m) \in F\left(\hat{R}\right)$. Since $F$ satisfies $\mu^{**}$, there exists a profile $\left(C_{\ell}\left(\hat{R}, g(m)\right)\right)_{\ell \in N}$ such that $C_{\ell}\left(\hat{R}, g(m)\right) \subseteq L\left(\hat{R}_{\ell}, g(m)\right) \cap Y$ for all $\ell \in N$. As $L\left(\hat{R}_i, g(m)\right) = L(R_i, g(m))$, $R_{-i} = \hat{R}_{-i}$, and $H = \{i\}$, Condition $\mu^{**}$(iv) implies that $g(m) \in F(R)$.

*Case 4*: $m$ corresponds to *Rule 3.*

Then, $g(m) \in \max_{R_{\ell}} Y$ for all $\ell \in N$. So, by Condition $\mu^{**}$(iii), $g(m) \in F(R)$. ■

**Proof of Theorem 4.** Let Assumption 1 hold and let $\mathcal{R}^n$ satisfy **RD**. Take any $F \in \mathcal{F}$ and let $\diamond \in N$ be an arbitrary agent index. Let $\gamma \equiv (M, g)$ be an $s$-mechanism.

*1. The necessity of Condition $M_s^{**}$.*

Suppose that $F$ is partially-honest implemented by $\gamma \equiv (M, g) \in \Gamma_{SP}$. From Theorem 3, it follows that $F$ satisfies Condition $M_s^{*}$. Furthermore, by using the same reasoning used in Theorem 2, it can readily be obtained that $F$ satisfies Condition $M_s^{**}$(iii) and Condition $M_s^{**}$(iv).

*2. The sufficiency of Condition $M_s^{**}$.*

Suppose that $F$ satisfies $M_s^{**}$. Then, for all $(R,x) \in \mathcal{R}^n \times X$ with $x \in F(R)$, $x \in Y$. Let $\gamma \equiv (M,g)$ be the mechanism defined in sub-section 3.1. For each $\ell \in N$, the set of truthful messages is that defined in (2). By construction, $\gamma \in \Gamma_{SP}$. Suppose that $R \in \mathcal{R}^n$ is the true state. The proof that $F(R) \subseteq NA\left(\gamma, \succsim^R\right)$ for any $H' \in \mathcal{H}$ can be given similar to the corresponding part in the proof of Theorem 2, so we omit it here. Conversely, to show that $NA\left(\gamma, \succsim^R\right) \subseteq F(R)$ for any $H' \in \mathcal{H}$, taking any $H \in \mathcal{H}$, let $m \in NE\left(\gamma, \succsim^R\right)$ for this $H$, and let $h$ be an arbitrary partially-honest agent in $H$. Let us consider the following cases.

*Case 1*: $m$ falls into *Rule 1.*

Then, $m$ is consistent with $x$ and $\bar{R} \in \mathcal{R}^n$, where $x \in F\left(\bar{R}\right)$. Thus, $g(m) = x$. Moreover, $C_\ell\left(\bar{R}_\ell, x\right) \subseteq L(R_\ell, x)$ for all $\ell \in N$. Suppose that $m_h \notin T_h^\gamma(R,F)$ for some $h \in H$. Suppose that $C_h\left(\bar{R}_h, x\right) = Y$. By changing her strategy $m_h$ into $m_h' \in T_h^\gamma(R,F)$, agent $h$ can trigger the modulo game and choose an agent index $k^h$ so that $\ell = \ell^*(m_h', m_{-h}) \neq h$. This implies $g(m_h', m_{-h}) = x$. Hence, $m \notin NE\left(\gamma, \succsim^R\right)$, a contradiction. Otherwise, let $C_h\left(\bar{R}_h, x\right) \neq Y$. By changing her strategy $m_h$ into $m_h' = (R_h, R_{h+1}, x, \diamond) \in T_h^\gamma(R,F)$, $(m_h', m_{-h})$ falls into *Rule 2* so that $g(m_h', m_{-h}) = x$. Then, $m \notin NE\left(\gamma, \succsim^R\right)$, a contradiction. Therefore, $m_h \in T_h^\gamma(R,F)$ for all $h \in H$. This reasoning is applied to any $H \in \mathcal{H}$, thus Condition $M_s^{**}$(i) implies $x \in F(R)$.

*Case 2*: $m$ falls into *Rule 2.*

Then, $m$ is $m_{-i}$ quasi-consistent with $\left(\bar{R}, x\right) \in \mathcal{R}^n \times Y$, where $x \in F\left(\bar{R}\right)$. Thus, $g(m) = x$. We proceed accordingly the following sub-cases: 1) $R_i^i \neq \bar{R}_i$ and $R_{i+1}^i \neq \bar{R}_{i+1}$ and 2) $R_i^i \neq \bar{R}_i$ and $R_{i+1}^i = \bar{R}_{i+1}$.[35]

*Sub-case 2.1.* $R_i^i \neq \bar{R}_i$ and $R_{i+1}^i \neq \bar{R}_{i+1}$.

So, $C_i\left(\bar{R}_i, x\right) \subseteq L(R_i, x)$ and $x \in \max_{R_\ell} Y$ for all $\ell \in N \backslash \{i\}$. By the definition of $g$, $m_h \in T_h^\gamma(R,F)$ for all $h \in H$; otherwise a contradiction can be obtained. Observe that if agent $i$ is a partially-honest agent, it must be the case that $R_i^{i-1} \neq R_i$ or $R_{i+1}^{i+1} \neq R_{i+1}$. To show this, suppose that $R_i^{i-1} = R_i$ and $R_{i+1}^{i+1} = R_{i+1}$. Then, agent $i \in H$ can change $m_i$ into $m_i' = (R_i, R_{i+1}, x, k^i) \in T_i^\gamma(R,F)$ and induce *Rule 1*. Then, $g(m_i', m_{-i}) = x$ and so $((m_i', m_{-i}), m) \in \succ_i^R$, which contradicts $m \in NE\left(\gamma, \succsim^R\right)$ for this $H$. Therefore, for any $H \in \mathcal{H}$, if $m \in NE\left(\gamma, \succsim^R\right)$ falls into *Rule 2* and $i \in H$, it has to be the case that $R_i^{i-1} \neq R_i$ or $R_{i+1}^{i+1} \neq R_{i+1}$. It follows that $i-1 \notin H$ or $i+1 \notin H$ if $i \in H$.

Suppose that $\#H > 1$. Condition $M_s^{**}$(ii.b) implies that $x \in F(R)$. Otherwise, let $\#H = 1$. If $H \subseteq N \backslash \{i\}$, Condition $M_s^{**}$(ii.b) implies that $x \in F(R)$. Finally, suppose that $H = \{i\}$. By following the same reasoning used in *Case 3* of the proof of Theorem 2, **RD**, Condition $M_s^{**}$(ii.a), and Condition $M_s^{**}$(iv) imply that $x \in F(R)$.

*Sub-case 2.2.* $R_i^i \neq \bar{R}_i$ and $R_{i+1}^i = \bar{R}_{i+1}$

Let $R_i^i = R_i'$ and $\bar{R}' \equiv \left(\bar{R}_{-i}, R_i'\right)$. We distinguish whether $x \in F\left(\bar{R}'\right)$ or not. Suppose that $x \notin F\left(\bar{R}'\right)$. Then, since $x \in F\left(\bar{R}\right)$, the same reasoning used above for *sub-case 2.1* carries over into this sub-case, so that $x \in F(R)$. Otherwise, let $x \in F\left(\bar{R}'\right)$. Then, there are two potential deviators, $i-1$ and $i$. Agent $\ell \in N \backslash \{i-1, i\}$ can attain any $y \in Y \backslash \{x\}$ by inducing *Rule 4*, so that $x \in \max_{R_\ell} Y$ as $m \in NE(\gamma, R)$. Consider agent $i-1$. Take any $y \in C_{i-1}\left(\bar{R}_{i-1}, x\right) = C_{i-1}\left(R_{i-1}^{i-2}, x\right)$. Suppose that $C_{i-1}\left(\bar{R}_{i-1}, x\right) \neq Y$. By changing $m_{i-1}$ to $m_{i-1}^* = \left(R_{i-1}^{i-1}, R_i^{i-1}, y, \diamond\right) \in M_{i-1}$, agent $i-1$ can obtain $y = g\left(m_{i-1}^*, m_{-(i-1)}\right)$ via *Rule 3*. In the case that $C_{i-1}\left(\bar{R}_{i-1}, x\right) = Y$, by changing $m_{i-1}$ to $m_{i-1}^* = \left(R_{i-1}^{i-1}, R_i^{i-1}, y, k^{i-1}\right) \in M_{i-1}$,

[35] The sub-case $R_i^i = \bar{R}_i$ and $R_{i+1}^i \neq \bar{R}_{i+1}$ is not explicitly considered as it can be proved similarly to the *sub-case 2.2* shown below.

agent $i-1$ can attain $y = g\left(m^*_{i-1}, m_{-(i-1)}\right)$ via *Rule 4* by an appropriate choice of $k^{i-1}$. It follows that $C_{i-1}\left(\bar{R}_{i-1}, x\right) \subseteq g\left(M_{i-1}, m_{-(i-1)}\right)$; then, $C_{i-1}\left(\bar{R}_{i-1}, x\right) \subseteq L\left(R_{i-1}, x\right)$ as $m \in NE\left(\gamma, R\right)$. As a similar argument applies to agent $i$, we have that $C_i\left(\bar{R}_i, x\right) \subseteq g\left(M_i, m_{-i}\right) \subseteq L\left(R_i, x\right)$ as $m \in NE\left(\gamma, R\right)$. Furthermore, by definition of $g$, $m_h \in T^{\gamma}_h\left(R, F\right)$ for all $h \in H$. Therefore, $x \in F\left(R\right)$ by $M^{**}_s$(i).

*Case 3*: $m$ falls into *Rule 3*.

Then, $m$ is $m_{-i}$ consistent with $x$ and $\bar{R} \in \mathcal{R}^n$, where $x \in F\left(\bar{R}\right)$. Moreover, $C_i\left(\bar{R}_i, x\right) \neq Y$. By the definition $g$ and $m \in NE\left(\gamma, \succcurlyeq^R\right)$, we have that $g\left(m\right) \in C_i\left(\bar{R}_i, x\right) \subseteq L\left(R_i, g\left(m\right)\right)$ and $g\left(m\right) \in \max_{R_\ell} Y$ for all $\ell \in N\backslash\left\{i\right\}$.[36] Moreover, $m_h \in T^{\gamma}_h\left(R, F\right)$ for all $h \in N$; otherwise a contradiction can be obtained. Suppose that $\#H > 1$. Condition $M^{**}_s$(ii.b) implies that $g\left(m\right) \in F\left(R\right)$. Otherwise, let $\#H = 1$. If $H \subseteq N\backslash\left\{i\right\}$, Condition $M^{**}_s$(ii.b) implies that $g\left(m\right) \in F\left(R\right)$. Finally, suppose that $H = \left\{i\right\}$. By following the same reasoning used in *Case 3* of Theorem 2, it follows from **RD**, Condition $M^{**}_s$(ii.a), and Condition $M^{**}_s$(iv) that $g\left(m\right) \in F\left(R\right)$.

*Case 4*: $m$ falls into *Rule 4*.

Then, $Y = g\left(M_\ell, m_{-\ell}\right)$ for all $\ell \in N$. As $m \in NE\left(\gamma, \succcurlyeq^R\right)$, $g\left(m\right) \in \max_{R_\ell} Y$ for all $\ell \in N$. Then, Condition $M^{**}_s$(iii) implies that $g\left(m\right) \in F\left(R\right)$. ■

**Proof of Theorem 6.** Let Assumption 1 hold and let $\mathcal{R}^n$ satisfy **RD**. Take any $F \in \mathcal{F}$ and let $\diamond \in N$ be an arbitrary agent index. Let $\gamma \equiv \left(M, g\right)$ be a self-relevant mechanism.

*1. The necessity of Condition $\lambda^{**}$.*

Suppose that $F$ is partially-honest implemented by $\gamma = \left(M, g\right) \in \Gamma_{SP}$. It is clear that $F$ satisfies Condition $\lambda^*$. Further, as in the proof of Theorem 2, we can see that $F$ satisfies Condition $\lambda^{**}$(iii) and Condition $\lambda^{**}$(v). Thus, $F$ satisfies Condition $\lambda^{**}$.

*2. The sufficiency of Condition $\lambda^{**}$.*

Let $F$ satisfy Condition $\lambda^{**}$. The proof can be obtained by using the self-relevant mechanism devised by Tatamitani (2001). We report it only for completeness. Define the outcome function $g : M \to X$ as follows. For all $m \in M$ and $\left(\bar{R}, x\right) \in \mathcal{R}^n \times Y$,

*Rule 1*: If $\left(R^\ell_\ell, x^\ell\right) = \left(\bar{R}_\ell, x\right)$ for all $\ell \in N$ and $x \in F\left(\bar{R}\right)$, then $g\left(m\right) = x$.
*Rule 2*: If there exists $i \in N$ such that $\left(R^\ell_\ell, x^\ell\right) = \left(\bar{R}_\ell, x\right) \neq \left(R^i_i, x^i\right)$ for all $\ell \in N\backslash\left\{i\right\}$, with $x^i \neq x \in F\left(\bar{R}\right)$, $i \in D\left(\left(R^i_i, \bar{R}_{-i}\right), x\right)$, and $C_i\left(\bar{R}_{-i}, x\right) \neq Y$, then

$$g\left(m\right) = \begin{cases} x^i & \text{if } \ x^i \in C_i\left(\bar{R}_{-i}, x\right) \\ x & \text{otherwise} \end{cases} .$$

*Rule 3*: If $\left(R^\ell_\ell, x^\ell\right) = \left(\bar{R}_\ell, x\right)$ for all $\ell \in N$, $x \notin F\left(\bar{R}\right)$, and $D\left(\bar{R}, x\right) \neq \varnothing$, then $g\left(m\right) = p\left(\bar{R}, x\right)$.
*Rule 4*: Otherwise, $g\left(m\right) = x^{\ell^*\left(m\right)}$ where $\ell^*\left(m\right) \equiv \sum\limits_{i \in N} k^i \left(\text{mod } n\right)$.[37]

For each $\ell \in N$, the set of truthful messages is that defined in (3). Moreover, by construction, $\gamma \in \Gamma_{SP}$. Observe that as $F$ satisfies $\lambda^{**}$, we have that for all $R \in \mathcal{R}^n$ and all $x \in F\left(R\right)$, $x \in Y$ holds. Take any $R \in \mathcal{R}^n$ and any $H \in \mathcal{H}$. The proof that $F\left(R\right) \subseteq NA\left(\gamma, \succcurlyeq^R\right)$ can be found in Tatamitani (2001), so we omit it here. To show that $NA\left(\gamma, \succcurlyeq^R\right) \subseteq F\left(R\right)$, let

[36] A detailed and exhaustive argument is provided in Lombardi and Yoshihara (2010).
[37] If the remainder is *zero*, the winner of the game is agent $n$.

$m \in NE\left(\gamma, \succcurlyeq^R\right)$ for this $H$, and let $h$ be an arbitrary partially-honest agent in $H$. Let us consider the following cases.

*Case 1*: $m$ falls into *Rule 1*.

Then, $m$ is such that for all $\ell \in N$, $m_\ell = \left(\bar{R}_\ell, x, \diamond\right)$ and $x \in F\left(\bar{R}\right)$. By definition of $g$ and the assumption that $m \in NE\left(\gamma, \succcurlyeq^R\right)$, we have that $C_\ell\left(R_{-\ell}, x\right) \subseteq L\left(R_\ell, x\right)$ for all $\ell \in N$. Suppose $m_h \notin T_h^\gamma\left(R, F\right)$ for some $h \in H$. Suppose that $C_h\left(\bar{R}_{-h}, x\right) = Y$. By changing her strategy $m_h$ to $m_h' = \left(R_h, x^h, k^h\right) \in T_h^\gamma\left(R, F\right)$ with $x^h \in Y \backslash \{x\}$, agent $h$ induces *Rule 4* and can obtain $\ell = \ell^*\left(m_{-h}, m_h'\right) \neq h$. Then, $g\left(m_h', m_{-h}\right) = x$. Hence, $\left(\left(m_{-h}, m_h'\right), m\right) \in \succ_h^R$, a contradiction. Otherwise, let $C_h\left(\bar{R}_{-h}, x\right) \neq Y$. By changing her strategy $m_h$ to $m_h' = \left(R_h, x^h, k^k\right) \in T_h^\gamma\left(R, F\right)$ with $x^h \in Y \backslash C_h\left(\bar{R}_{-h}, x\right)$, $\left(m_h', m_{-h}\right)$ falls into *Rule 2* as $h \in D\left(\bar{R}_{-h}, R_h\right)$. Then, $g\left(m_h', m_{-h}\right) = x$. Again, $\left(\left(m_h', m_{-h}\right), m\right) \in \succ_h^R$, a contradiction. We conclude that $m_h \in T_h^\gamma\left(R, F\right)$ for all $h \in H$. Condition $\lambda^{**}$(i) implies $x \in F\left(R\right)$.

*Case 2*: $m$ falls into *Rule 2*.

Then, $m$ is such that $m_\ell = \left(\bar{R}_\ell, x, \diamond\right)$ for any $\ell \in N \backslash \{i\}$ and $m_i = \left(R_i^i, x^i, \diamond\right)$, with $x^i \neq x$, $i \in D\left(\left(\bar{R}_{-i}, R_i^i\right), x\right)$, and $C_i\left(\bar{R}_{-i}, x\right) \neq Y$. By the definition of $g$, we have that $C_i\left(\bar{R}_{-i}, x\right) \subseteq g\left(M_i, m_{-i}\right)$.

Next, we claim that $g\left(M_\ell, m_{-\ell}\right) = Y$ for all $\ell \in N \backslash \{i\}$. We proceed according to whether $\#Y = 2$ and $n = 3$ or not.

*Sub-case 2.1. not*[$\#Y = 2$ and $n = 3$]

Suppose that $\#Y > 2$. Take any $\ell \in N \backslash \{i\}$. Then, agent $\ell$ can induce the modulo game by choosing any $y \in Y \backslash \{x, x^i\}$ and changing $m_\ell$ into $m_\ell^* = \left(\bar{R}_\ell, y, k^\ell\right)$. To attain $y$, agent $\ell$ has only to adjust $k^\ell$ by which $\ell^*\left(m_\ell^*, m_{-\ell}\right) = \ell$. To attain $x$ (resp., $x^i$), agent $\ell$ has only to adjust $k^\ell$ by which $\ell^*\left(m_\ell^*, m_{-\ell}\right) = j$ for $j \in N \backslash \{\ell, i\}$ (resp., $\ell^*\left(m_\ell^*, m_{-\ell}\right) = i$). Therefore, $Y \subseteq g\left(M_\ell, m_{-\ell}\right)$ for any $\ell \in N \backslash \{i\}$. Otherwise, let $\#Y = 2$. Then, $n > 3$. Take any $\ell \in N \backslash \{i\}$. Choosing $x^\ell = x^i$, agent $\ell$ can make $\#\left\{\ell \in N | x^\ell = x\right\} \geq 2$ and $\#\left\{\ell \in N | x^\ell \neq x\right\} \geq 2$. As the outcome is determined by *Rule 4*, agent $\ell$ can attain any outcome in $Y$ by appropriately choosing $k^\ell$. Therefore, $Y \subseteq g\left(M_\ell, m_{-\ell}\right)$ for any $\ell \in N \backslash \{i\}$.

*Sub-case 2.2.* $\#Y = 2$ and $n = 3$

Then, $Y = \{x, x^i\}$ and $N = \{i, \ell, \ell'\}$. Moreover, $g\left(m\right) = x$. Agent $\ell$ can change her strategy $m_\ell$ to $m_\ell^* = \left(\bar{R}_\ell, x^i, k^\ell\right)$. If $\ell' \notin D\left(\bar{R}, x^i\right)$, then the outcome is determined by *Rule 4*. Therefore, agent $\ell$ can attain $x^i \in Y$ by appropriately choosing the integer index $k^\ell$. Otherwise, let $\ell' \in D\left(\bar{R}, x^i\right)$. Suppose that $C_{\ell'}\left(\bar{R}_{-\ell'}, x^i\right) \neq Y$. As $\left(m_\ell^*, m_{-\ell}\right)$ falls into *Rule 2*, it follows that $g\left(m_i^*, m_{-i}\right) = x^i$. Otherwise, agent $\ell$ can attain $x^i$ by appropriately choosing $k^\ell$, since the outcome is determined by *Rule 4*.

Since $m \in NE\left(\gamma, \succcurlyeq^R\right)$, we have that $C_i\left(\bar{R}_{-i}, x\right) \subseteq L\left(R_i, g\left(m\right)\right)$ and $g\left(m\right) \in \max_{R_\ell} Y$ for all $\ell \in N \backslash \{i\}$. Moreover, by the definition of $g$, we have that $m_h \in T_h^\gamma\left(R, F\right)$ for all $h \in H$; otherwise $m \in NE\left(g, \succcurlyeq^R\right)$ cannot hold, a contradiction (details available from the authors). Suppose that $\#H > 1$, Condition $\lambda^{**}$(ii.b) implies that $g\left(m\right) \in F\left(R\right)$. Otherwise, let $\#H = 1$. Suppose that $H \subseteq N \backslash \{i\}$. Again, Condition $\lambda^{**}$(ii.b) implies that $g\left(m\right) \in F\left(R\right)$. Finally, let $H = \{i\}$. By following the same reasoning used in *Case 3* of Theorem 2, it follows from **RD**, Condition $\lambda^{**}$(ii.a), and Condition $\lambda^{**}$(v) that $g\left(m\right) \in F\left(R\right)$.

*Case 3*: $m$ falls into *Rule 3*.

Then, $m$ is such that $m_\ell = \left(R_\ell^\ell, x^\ell, \diamond\right) = \left(\bar{R}_\ell, x, \diamond\right) \in M_\ell$, $x \notin F\left(\bar{R}\right)$, and $D\left(\bar{R}, x\right) \neq \varnothing$. By *Rule 3*, $g\left(m\right) = p\left(\bar{R}, x\right)$.

Take any $i \in D(\bar{R}, x)$. We show that $C_i(\bar{R}_{-i}, x) \subseteq g(M_i, m_{-i})$. As $i \in D(\bar{R}, x)$, there exists $R'_i \in \mathcal{R}_i$ such that $x \in F(R'_i, \bar{R}_{-i})$. By changing $m_i$ to $m'_i = (R'_i, x, \diamond)$, agent $i$ can induce *Rule 1* and obtain $g(m'_i, m_{-i}) = x$. Take any $x^i \in C_i(\bar{R}_{-i}, x) \setminus \{x\}$. Suppose $C_i(\bar{R}_{-i}, x) \neq Y$. By changing $m_i$ to $m'_i = (\bar{R}_i, x^i, k^i)$, agent $i$ can induce *Rule 2* and obtain $g(m'_i, m_{-i}) = x^i$. Suppose $C_i(\bar{R}_{-i}, x) = Y$. Then, the modulo game is triggered and agent $i$ can attain $x^i$ by choosing $k^i$ appropriately.

Take any $\ell \in N \backslash D(\bar{R}, x)$. We show that $Y \subseteq g(M_\ell, m_{-\ell})$. Then, agent $\ell$ can induce the modulo game by choosing any $x^\ell \in Y \setminus \{x\}$ and changing $m_\ell$ into $m'_\ell = (\bar{R}_\ell, x^\ell, k^\ell)$. To attain $x$ and $x^\ell$, agent $\ell$ has only to adjust $k^\ell$ by which $\ell^*(m'_\ell, m_{-\ell}) = i$ for $i \in N \setminus \{\ell\}$ and $\ell^*(m^*_\ell, m_{-\ell}) = \ell$, respectively.

Since $m \in NE(\gamma, \succcurlyeq^R)$, $C_i(\bar{R}_{-i}, x) \subseteq L(R_i, p(\bar{R}, x))$ holds for any $i \in D(\bar{R}, x)$, and $p(\bar{R}, x) \in \max_{R_\ell} Y$ holds for all $\ell \in N \backslash D(\bar{R}, x)$. Moreover, by the definition of $g$, we have that $m_h \in T_h^\gamma(R, F)$ for any $h \in H$; otherwise $m \in NE(g, \succcurlyeq^R)$ cannot hold, a contradiction (details available from the authors). Condition $\lambda^{**}$(iv) implies that $p(\bar{R}, x) \in F(R)$.

*Case 4*: $m$ falls into *Rule 4*.

Then, $g(m) = x^{\ell^*(m)}$ where $\ell^*(m) \in N$ is the winner of the modulo game. We show that $Y \subseteq g(M_\ell, m_{-\ell})$ for any $\ell \in N$. Take any $i \in N$ and consider the following two sub-cases. Let $(R_\ell^\ell)_{\ell \in N} \equiv \bar{R}$.

*Sub-case 4.1*: For all $\ell, \ell' \in N \setminus \{i\}$, $x^\ell = x^{\ell'}$.

Suppose that $x^\ell = x$ for all $\ell \in N$. As $m$ falls into *Rule 4*, it follows that $x \notin F(\bar{R})$ and $D(\bar{R}, x) = \varnothing$. By changing $m_i$ to $m'_i = (\bar{R}_i, x^i, k^i)$ with $x^i \in Y$, agent $i$ can trigger the modulo game and obtain $g(m'_i, m_{-i})$ by appropriately choosing $k^i$. Therefore, $Y \subseteq g(M_i, m_{-i})$. On the other hand, suppose that $x^\ell = x$ for all $\ell \in N \setminus \{i\}$, and $x^i \neq x$. Take any $\hat{x}^i \in Y \setminus \{x\}$. Since $g(m) = x^{\ell^*(m)}$, it follows that either $i \notin D(\bar{R}, x)$ or $i \in D(\bar{R}, x)$, and $C_i(\bar{R}_{-i}, x) = Y$. Therefore, by deviating from $m_i$ to $m'_i = (\bar{R}_i, \hat{x}^i, k^i)$, agent $i$ can trigger the modulo game. To attain $x$ and $\hat{x}^i$, agent $i$ has only to adjust $k^i$ so that $\ell^*((m'_i, m_{-i})) \in N \setminus \{i\}$ and $\ell^*((m'_i, m_{-i})) = i$, respectively. Again, we have that $Y \subseteq g(M_i, m_{-i})$.

*Sub-case 4.2*: For some $\ell, \ell' \in N \setminus \{i\}$, $x^\ell \neq x^{\ell'}$.

Suppose that $\#Y = 2$ and let $Y = \{x^\ell, x^{\ell'}\}$. By changing $m_i$ to $m'_i = (\bar{R}_i, x^i, k^i)$, agent $i$ induces the modulo game. To attain $x^\ell$ and $x^{\ell'}$, agent $i$ has only to adjust the integer index $k^i$ so that $\ell^*((m'_i, m_{-i})) = \ell$ and $\ell^*((m'_i, m_{-i})) = \ell'$, respectively. Otherwise, let $\#Y > 2$. Take any $x^i \in Y \setminus \{x^\ell, x^{\ell'}\}$. By deviating from $m_i$ to $m'_i = (\bar{R}_i, x^i, k^i)$, agent $i$ can trigger the modulo game. To attain $x^\ell$, $x^{\ell'}$, and $x^i$, agent $i$ has only to adjust $k^i$ so that $\ell^*((m'_i, m_{-i})) = \ell$, $\ell^*((m'_i, m_{-i})) = \ell'$, and $\ell^*((m'_i, m_{-i})) = i$, respectively. Again, we have that $Y \subseteq g(M_i, m_{-i})$.

Since $Y \subseteq g(M_i, m_{-i})$ for any $i \in N$ and $m \in NE(\gamma, \succcurlyeq^R)$, we have that $g(m) \in \max_{R_\ell} Y$ for all $\ell \in N$. Condition $\lambda^{**}$(iii) implies that $g(m) \in F(R)$. ■

**Proof of Theorem 8.** Let Assumption 1 and **RD** hold. Let $\mathcal{H} = \{\{1\}, \{2\}\}$. Take any $F \in \mathcal{F}$ defined on $\mathcal{R}^2$. Let $\gamma \equiv (M, g)$ be a mechanism. Let $h$ denote the unique partially-honest agent in $N$.

*1. The necessity of Condition $\mu 2^{**}$.*

Let $F$ be partially-honest implemented by $\gamma \in \Gamma_{SP}$. Let $Y \equiv g(M)$. Take any $H \in \mathcal{H}$, any $R \in \mathcal{R}^2$, and any $x \in F(R)$. Then, there exists an $m(R, x) \in NE(\gamma, \succcurlyeq^R)$ for this $H$ such that $g(m(R, x)) = x$. Observe that $m_h(R, x) \in T_h^\gamma(R, F)$ as $\gamma \in \Gamma_{SP}$. For all $\ell \in N$, let $C_\ell(R, x) \equiv g(M_\ell, m_i(R, x))$, where $i \in N \setminus \{\ell\}$. Then, $g(M_\ell, m_i(R, x)) \subseteq L(R_\ell, x) \cap Y$

for all $\ell \in N$. From Theorem 2, it follows that $F$ satisfies Conditions $\mu^{**}$. Next, we show that $F$ satisfies Condition $\mu2^{**}$(v). Pick any $(x', R') \in X \times \mathcal{R}^2$ with $x' \in F(R')$, and take any $R^* \in \mathcal{R}^2$. Since $x' \in F(R')$, it follows that there exists an $m(R', x') \in NE\left(\gamma, \succcurlyeq^{R'}\right)$ and $g(m(R', x')) = x'$, where $m_h(R', x') \in T_h^{\gamma}(R', F)$. Let $e \equiv e(x', R', x, R) = g(m_1(R, x), m_2(R', x'))$. Then, defining $C_1(R', x') \equiv g(M_1, m_2(R', x'))$ and $C_2(R, x) \equiv g(m_1(R, x), M_2)$, $e \in C_1(R', x') \cap C_2(R, x)$ holds. Thus, $F$ satisfies $\mu2^{**}$(v.a). It is also clear that $F$ meets Condition $\mu2^{**}$(v.c), since $R = R' = R^*$ implies that every agent is truthful and $e$ is optimal at state $R^*$. Finally, we check $\mu2^{**}$(v.b). Let $x \neq x'$ and $R \neq R'$. Moreover, suppose that $C_1(R', x') \subseteq L(R_1^*, e)$, $C_2(R, x) \subseteq L(R_2^*, e)$, and $e \notin F(R^*)$. Suppose that $R = R^*$. Assume, to the contrary, that $H = \{1\}$. Then, $m_1(R, x) \in T_1^{\gamma}(R^*, F)$. Since there cannot be any profitable deviation, we have that $e \in NA\left(\gamma, \succcurlyeq^{R^*}\right)$ for $H = \{1\}$, a contradiction. Thus, $H = \{2\}$. Similarly, we obtain $H = \{1\}$ if $R' = R^*$. In summary, $F$ satisfies Condition $\mu2^{**}$(v).

*2. The sufficiency of Condition $\mu2^{**}$.*

Suppose that $F$ satisfies Condition $\mu2^{**}$. Then, $F(\mathcal{R}^2) \subseteq Y$. For each $\ell \in N$, define $M_\ell$ as follows

$$M_\ell \equiv \left\{m_\ell = \left(R^\ell, x^\ell, y^\ell, k^\ell\right) \in \mathcal{R}^2 \times X \times Y \times \mathbb{Z}_+ \mid x^\ell \in F\left(R^\ell\right)\right\}, \tag{4}$$

where $\mathbb{Z}_+$ is the set of nonnegative integers. The set of truthful messages is that defined in (1).

Define the outcome function $g : M \to X$ as follows: For all $m \in M$, for $i, j \in N$, with $i \neq j$:

*Rule 1:* If $(R^i, x^i) = (R^j, x^j)$ and $k^i = k^j = 0$, then $g(m) = x^i$.
*Rule 2:* If $k^i > k^j = 0$, then

$$g(m) = \begin{cases} y^i & \text{if } \; y^i \in C_i(R^j, x^j) \\ e \equiv e(x^2, R^2, x^1, R^1) & \text{otherwise.} \end{cases}$$

*Rule 3:* If $(R^1, x^1) \neq (R^2, x^2)$ and $k^1 = k^2 = 0$, then

$$g(m) = \begin{cases} x^1 & \text{if } \; x^1 = x^2 \\ e \equiv e(x^2, R^2, x^1, R^1) & \text{otherwise.} \end{cases}$$

*Rule 4:* If $k^1 \geq k^2 > 0$, then, $g(m) = y^1$.
*Rule 5:* Otherwise, $g(m) = y^2$.

The outcome $e \equiv e(x^2, R^2, x^1, R^1)$ is the outcome specified in Condition $\mu2^{**}$(v.a). Observe that $\gamma \in \Gamma_{SP}$, by construction.

Suppose that $R \in \mathcal{R}^2$ is the true state and take any $H \in \mathcal{H}$.

Let $x \in F(R)$ and suppose that for all $\ell \in N$, $m_\ell(R, x) = (R, x, x, 0) \in T_\ell^{\gamma}(R, F)$. *Rule 1* implies that $g(m) = x$. By the definition of $g$, any deviation by agent $\ell \in N$ leads to an outcome in $C_\ell(R, x)$, so that $g(M_\ell, m_i(R, x)) = C_\ell(R, x)$, where $i \in N \setminus \{\ell\}$. Since $C_\ell(R, x) \subseteq L(R_\ell, x)$, such deviations are not profitable. It follows that $x \in NA\left(\gamma, \succcurlyeq^R\right)$ for this $H$. To show that $NA\left(\gamma, \succcurlyeq^R\right) \subseteq F(R)$, let $m \in NE\left(\gamma, \succcurlyeq^R\right)$ and let us consider the following cases.

*Case 1: $m$ corresponds to Rule 1.*

Suppose that $m$ falls into *Rule 1*. Then, $g(m) = x^1$. By the definition of $g$, it follows that $m_h \in T_h^{\gamma}(R, F)$. Then, $x^1 = x^2 \in F(R)$.

*Case 2:* $m$ corresponds to *Rule 2*.

Without loss of generality, let $i = 1$. Then, $g(m_1, M_2) = Y \subseteq L(R_2, g(m))$ and $C_1(R^2, x^2) = g(M_1, m_2) \subseteq L(R_1, g(m))$. By the definition of $g$, $m_h \in T_h^\gamma(R, F)$. Suppose that $H = \{2\}$. Condition $\mu 2^{**}$(ii.c) implies that $g(m) \in F(R)$ as $R^2 = R$. Otherwise, let $H = \{1\}$. Following the same reasoning used in *Case 3* of Theorem 2, it follows from **RD**, Condition $\mu 2^{**}$(ii.a), and Condition $\mu 2^{**}$(iv) that $g(m) \in F(R)$.

*Case 3:* $m$ corresponds to *Rule 3*.

Then, $C_1(R^2, x^2) = g(M_1, m_2(R^2, x^2)) \subseteq L(R_1, g(m))$ and $C_2(R^1, x^1) = g(m_1(R^1, x^1), M_2) \subseteq L(R_2, g(m))$. Observe that $m_h(R^h, x^h) \in T_h^\gamma(R, F)$. If $x^1 = x^2$, then $g(m) \in F(R)$. Otherwise, let $x^1 \neq x^2$. Suppose that $R^1 = R^2$. Then, since $F$ satisfies Condition $\mu 2^{**}$, it follows that $(g(m), x^2) \in I_1$ and $(g(m), x^1) \in I_2$. Condition $\mu 2^{**}$(v.c) implies that $g(m) \in F(R)$. Finally, let $R^1 \neq R^2$. Suppose that $H = \{1\}$, so that $R^1 = R$. Condition $\mu 2^{**}$(v.b.1) implies that $g(m) \in F(R)$. Otherwise, let $H = \{2\}$, and so $R^2 = R$. Condition $\mu 2^{**}$(v.b.2) implies that $g(m) \in F(R)$.

*Cases 4:* $m$ corresponds to *Rule 5* or *Rule 6*.

Then, $g(M_1, m_2) = Y \subseteq L(R_1, g(m))$ and $g(m_1, M_2) = Y \subseteq L(R_2, g(m))$. Condition $\mu 2^{**}$(iii) implies that $g(m) \in F(R)$. ■

**Proof of Theorem 10.** Let Assumption 1 and **RD** hold. Take any $F \in \mathcal{F}$ defined on $\mathcal{R}^2$. Let $\gamma \equiv (M, g)$ be a mechanism.

*1. The necessity of Condition* $\mu 2^{\circ\circ}$.

Suppose that $F$ is partially-honest implemented by $\gamma \in \Gamma_{StP}$. From Theorem 8, Condition $\mu 2^{**}$ is satisfied. Furthermore, as it is clear that $F$ satisfies Condition $\mu 2^{\circ\circ}$(vi), we conclude that $F$ meets Condition $\mu 2^{\circ\circ}$.

*2. The sufficiency of Condition* $\mu 2^{\circ\circ}$.

Suppose that $F$ satisfies Condition $\mu 2^{\circ\circ}$. Then, $F(\mathcal{R}^2) \subseteq Y$. Consider the mechanism $\gamma$ constructed in Theorem 8. Clearly, $\gamma \in \Gamma_{StP}$. Moreover, let the set of truthful messages be that defined in (1).

Suppose that $R \in \mathcal{R}^2$ is the true state and pick any $H \in \mathcal{H}$. The proof that $F(R) \subseteq NA(\gamma, \succcurlyeq^R)$ follows from Theorem 8. Then, to show that $NA(\gamma, \succcurlyeq^R) \subseteq F(R)$ for this $H$, let $m \in NE(\gamma, \succcurlyeq^R)$ for this $H$. As in Theorem 8, we have to consider several cases. The proof that $g(m) \in F(R)$ follows from the same arguments used in Theorem 8 whenever *Rule 1*, *Rule 3*, *Rule 4*, or *Rule 5* applies to $m$. Therefore, suppose that $m$ falls into *Rule 2*. Without loss of generality, let $i = 1$. Then, $g(m_1, M_2) = Y \subseteq L(R_2, g(m))$ and $C_1(R^2, x^2) \subseteq g(M_1, m_2) \subseteq L(R_1, g(m))$. By the definition of $g$, we have that $m_h \in T_h^\gamma(R, F)$ for all $h \in H$. Suppose that $\#H = 1$. Then, $g(m) \in F(R)$ by *Case 2* of Theorem 8. Suppose that $\#H = 2$. Then, Condition $\mu 2^{\circ\circ}$(vi) implies that $g(m) \in F(R)$, as sought. ■