Implementation with Near-Complete Information: The Case of Subgame Perfection*

Takashi Kunimoto† and Olivier Tercieux ‡

First Draft: February 2009
This Draft: August 2009

Abstract

While monotonicity is a necessary and almost sufficient condition for Nash implementation and often a demanding one, almost any (non-monotonic, for instance) social choice rule can be implemented using undominated Nash or subgame perfect equilibrium. By requiring solution concepts to have closed graph in the limit of complete information, Chung and Ely (2003) show that only monotonic social choice rules can be implemented in the closure of the undominated Nash equilibrium correspondence. In this paper, we show that only monotonic social choice rules can be implemented in the closure of the subgame perfect equilibrium/sequential equilibrium correspondence. Our robustness result helps understand the limits of subgame pefect implementation, which is widely used in applications. We discuss the implications of our result for the literature on incomplete contracts.

Journal of Economic Literature Classification Numbers: C79, D82

Keywords: Monotonicity, Subgame Perfect Implementation, Robust implementation

*This research began when Olivier Tercieux was Member of the Institute for Advanced Study at Princeton. He gratefully acknowledges the Institute's hospitality and financial support through the Deutsche Bank membership. Takashi Kunimoto gratefully acknowledges financial supports from FQRSC and SSHRC of Canada. This paper benefits from the audiences at Kobe, McGill, 2009 Canadian Economic Theory Conference in Toronto, 2009 Game Theory Workshop at Hitotsubashi, 2008 NSF/NBER conference on general equilibrium and mathematical economics at Brown. Helpful conversations and comments from Philippe Aghion, Drew Fudenberg, Ed Green, Oliver Hart, Philippe Jehiel, Hitoshi Matsushima, Eiichi Miyagawa are gratefully acknowledged. All remaining errors are, of course, solely by the authors.

†Department of Economics, McGill University and CIREQ, Montreal, Quebec, Canada; takashi.kunimoto@mcgill.ca

‡Paris School of Economics, Paris, France; tercieux@pse.ens.fr

1 Introduction

Suppose that the society has a social choice rule which associates with each environment a subset of possible outcomes. The theory of implementation is concerned with characterizing the relationship between the structure of the institution (or mechanism) through which individuals interact and the outcome of that interaction, given a social choice rule and a domain of environments.

Maskin (1999) shows a condition called *monotonicity* is necessary and almost sufficient for Nash implementation. It turns out that monotonicity is quite a demanding condition and the literature tried to obtain less restrictive characterizations using *refinements* of Nash equilibrium. Using subgame perfect equilibrium, Moore and Repullo (1988) dispense with monotonicity and provide a sufficient condition for subgame perfect implementation.[1] As a different refinement, Palfrey and Srivastva (1991) propose *undominated Nash equilibrium* and prove that almost any social choice rule is implementable in undominated Nash equilibrium. Therefore, allowing for the use of refinements of Nash equilibrium, one can significantly expand the class of implementable social choice rules.

Chung and Ely (2003) investigate the robustness of undominated Nash implementation to incomplete information.[2] In so doing, they require that solution concepts have closed graph in the limit of complete information. Then, Chung and Ely (2003) conclude that when preferences are strict (or more generally hedonic), only monotonic social choice rules can be implemented in the closure of the undominated Nash equilibrium correspondence. Following the approach by Chung and Ely (2003), this paper investigates the robustness of *any* subgame perfect implementing mechanism to incomplete information. We show that only monotonic social choice rules can be implemented in the closure of the subgame perfect/sequential equilibrium correspondence. Hence, our result implies that there might be little difference between sequential mechanisms and static mechanisms, once we insist on robustness. This is due to the fact that a small amount of incomplete information expands the set of consistent beliefs of players along the game tree and so allows to

[1] Abreu and Sen (1990) further refine the analysis of Moore and Repullo (1988) and obtain a necessary and almost sufficient condition for subgame perfect implementation. Finally, Vartiainen (2007) obtains a full characterization.

[2] The type of perturbation used in Chung and Ely (2003) weakens common knowledge into common p-belief with p close to 1. Common p-belief is introduced in Monderer and Samet (1989). This is a "smaller"perturbation and less demanding than the one used for instance in Oury and Tercieux (2009). See also Kunimoto (2008) for a characterization of the perturbation used in this paper.

sustain additional sequential equilibria. The failure of monotonicity allows us to turn a "bad" Nash outcome into a "bad" sequential equilibrium outcome. Related observations have been made in the game theory literature by Fudenberg, Kreps, and Levine (1988). While similar in spirit, we make a very distinct argument. In this paper, we fix the payoff space and perturb only agents' beliefs over the fixed payoff space. This guarantees that the set of messages in the mechanism remains cheap-talk. Fudenberg, Kreps, and Levine (1988), on the other hand, are concerned with the situation in which the set of payoff states is not common knowledge, i.e., there are "crazy" types.

We put our result in a broader perspective. Since the early works of Grossman and Hart (1986) and Hart and Moore (1990), the incomplete contracts literature often cites indescribable contingencies as a major obstacle to the creation of complete contracts. Maskin and Tirole (1999a,b), however, argue that the literature's justification for incomplete contracts is conceptually problematic. Using the agents' minimum foresight concerning the possible payoff contingencies, they show that the inability to describe future contingencies by itself places no constraints on contracting. This is the so-called irrelevance theorem. To show this, Maskin and Tirole (1999a) reduce their task to checking sufficient conditions for subgame perfect implementation. Then, our result enables us to assess the robustness of Maskin and Tirole's irrelevance theorem. In fact, we can conclude that their implementing mechanism is not robust because a small amount of incomplete information necessitates that we should focus only on monotonic social choice rules. The paper by Moore and Repullo (1988) had a large impact and it is not difficult to find many other applications of subgame perfect implementation in the literature. For instance, Miyagawa (2002) shows that while many axiomatic bargaining solutions are not monotonic, they can be implemented in subgame perfect equilibrium by a four-stage sequential mechanism. Miyagawa's (2002) mechanism to implement bargaining solutions cannot also escape from our robustness argument.

There is a related paper by Aghion, Fudenberg, and Holden (henceforth, AFH) (2009). They also consider the question of subgame perfect implementation with almost complete information. AFH (2009) focus on a special class of mechanisms (henceforth, the Moore-Repullo mechanism) in the spirit of the one defined in Section 5 of Moore and Repullo (1988).[3] Under the assumption of complete information, given any social choice rule,

[3] In Section 4.1 of their paper, AFH (2009) go beyond the Moore-Repullo mechanism and obtain the same conclusion in mechanisms satisfying the following properties:(1) there are three stages; (2) there are

the Moore-Repullo mechanism guarantees that telling the truth is the unique subgame perfect equilibrium. In the same Moore-Repullo mechanism, however, AFH (2009) exhibit *some* social choice rules where telling the truth is not an (sequential) equilibrium when introducing a small amount of incomplete information. Loosely speaking, AFH (2009) stress the failure of the lower hemi-continuity of the equilibrium correspondence in the limit of complete information. On the contrary, our paper shows that the introduction of a small amount of incomplete information may induce new "bad" equilibria, i.e. equilibria that do not implement. This corresponds to the failure of the upper hemi-continuity of the equilibrium correspondence in the limit of complete information. When considering implementation problems, we believe that this is a meaningful requirement that indeed follows previous approaches (see Chung and Ely (2003)). While the motivation in AFH (2009) is similar to the present paper in spirit, our "robustness tests" are different and the results are also very different: (1) Our result is mechanism-free: we do not consider a fixed mechanism but a very general class of mechanisms that contains the one studied by AFH (2009); (2) our non-robustness result applies to any social choice rule that is not monotonic, while AFH (2009) focus on some social choice functions that fails their robustness test. Thus, we are able to relate the non-robust feature of any subgame perfect implementing mechanism to non-monotonicity of social choice rules.

The rest of the paper is organized as follows: In Section 2 we introduce the preliminary notation and definitions. Section 3 defines robust subgame perfect implementation. In Section 4, we state the main theorem and illustrate the main idea of this paper through an example. Section 5 concludes and illustrates the implications of our result for the incomplete contract literature through a hold-up problem. Finally, the proof of our main theorem is provided in the Appendix.

2 Setting

There is a finite set $N = \{1, ..., n\}$ of players, and a set A of social alternatives, or outcomes. There is a finite set Θ of states of nature. Associated with each state θ is a preference profile $\succeq^\theta$ which is a list $(\succeq_1^\theta, ..., \succeq_n^\theta)$. Players do not observe the state directly, but are informed of the state via signals. Player i's signal set is S_i which, for simplicity, we identify with Θ. A

two equally likely states of nature; (3) only one agent moves at each stage; and (4) pure strategies are only considered.

signal profile is an element $s = (s_1, ..., s_n) \in S \equiv \times_{i \in N} S_i$. When the realized signal profile is s, each player i observes only his own signal s_i. We let μ denote the prior probability over $\Theta \times S$, and let $\mathcal{P}$ be the set of all such priors. We note $\mu(\cdot \mid s_i)$ for the probability measure over $\Theta \times S$ conditional on s_i. Let s^θ be the signal profile in which each player's signal is s_i^θ. *Complete information* refers to the environments in which $\mu(\theta, s) = 0$ whenever $s \neq s^\theta$ (μ will be then referred to as a complete information prior). Under complete information, the state, and hence the full profile of preferences is always common knowledge among agents. We will assume for each i and $\theta : \mu(s_i^\theta) \equiv [\text{marg}_{S_i}\mu](s_i^\theta) > 0$ so that Bayes rule is well-defined. Given a prior μ over $\Theta \times S$, we will sometimes abuse notations and write $\mu(\theta)$ for $[\text{marg}_\Theta \mu](\theta)$. Besides, given $s_{-i} \in S_{-i}$, we will also write $\mu(s_{-i})$ as $[\text{marg}_{S_{-i}}\mu](s_{-i})$. Finally, given some arbitrary countable space X, δ_x will denote the probability measure that puts probability 1 on $\{x\} \subset X$.

A *social choice correspondence* (SCC) is a mapping $\mathcal{F}$ which associates a subset of A with each $\theta \in \Theta$. A single-valued social choice correspondence is a social choice *function* (SCF) denoted f. Hence, any selection of SCC $\mathcal{F}$ is a social choice function. A *mechanism* is an extensive game form $\Gamma = (\mathcal{H}, M, g)$ where $\mathcal{H}$ is a set of histories h. $M = M_1 \times \cdots \times M_n$ and $M_i = \times_{h \in \mathcal{H}} M_i(h)$ for all i. An element of $M(h) = M_1(h) \times \cdots \times M_n(h)$, say $m(h) = (m_1(h), ..., m_n(h))$ is a message profile at h while $m_i(h)$ is i's message at h. If $\#M_i(h) > 1$ and $\#M_j(h) > 1$ then agents i and j move simultaneously after history h, whereas if $\#M_i(h) > 1$ and $\#M_j(h) = 1$ for all $j \neq i$ then agent i is the only one to move. Histories and messages are tied together by the property that $M(h) = \{m : (h, m) \in \mathcal{H}\}$. An element of M_i is a pure strategy; and an element of M is a pure strategy profile. We sometimes write $m \mid_h = (m_1 \mid_h, ..., m_n \mid_h)$ for the profile of pure strategies starting from history h.

There is an initial history $\emptyset \in \mathcal{H}$, and each history h_t is represented by a sequence with finite length $t : (\emptyset, m^1, m^2, ..., m^{t-1}) = h_t$ where for each $k : m^k \in M(h_k)$.[4] If for $t' \geq t+1 : h_{t'} = (h_t, m^t, ..., m^{t'-1})$, then $h_{t'}$ *follows* history h_t. As Γ contains finitely many stages, there is a set of terminal histories[5] $H_T \subset \mathcal{H}$ such that $H_T = \{h \in \mathcal{H}$:there is no h' following $h\}$. Given any strategy profile m and any history h, there is a unique terminal

[4]As Moore and Repullo (1988), we restrict ourselves to mechanisms with finitely many stages. We allow agents to move simultaneously at some nodes, so mechanisms need not be with perfect information. However, at each node, all agents are assumed to know the entire history of the play.

[5]Note that $M(h) = \{m : (h, m) \in \mathcal{H}\} = \emptyset$ for any $h \in H_T$.

history denoted $h_T[m,h]$. Formally, let $\mathcal{Z}: M \times \mathcal{H} \to \mathcal{H}$ be the mapping where

$$\mathcal{Z}[m,h] = \begin{cases} (h, m(h)) \text{ if } h \notin H_T \\ h \text{ otherwise} \end{cases}$$

is the history that immediately follows h whenever possible given that strategy profile m has been played; and so $h_T[m,h] = \lim_{k\to\infty} \mathcal{Z}^k[m,h]$ where $\mathcal{Z}^k[m,h] = \mathcal{Z}[m, \mathcal{Z}^{k-1}[m,h]]$. Finally, the *outcome function* $g : H_T \to A$ specifies an outcome for each terminal history. We will also note $g(m;h)$ for the outcome that obtains when agents use strategy profile m starting from history h i.e. $g(m;h) = g(h_T[m,h])$.

Assumption 1 *$M_i(h)$ is countable for each i and h.*

Remark: This assumption is useful when using sequential equilibrium and avoids technical complications due to the use of measures over uncountable spaces. We, however, do not believe that our results critically depend on the countability assumption. We refer the reader to Duggan (1997) for the treatment of general (uncountable) message spaces. In addition, in our setting where the set of states has been assumed to be finite, the famous mechanism by Moore and Repullo (1988, Section 5) uses only a finite set of messages.

A stage mechanism Γ together with a profile $\theta \in \Theta$ defines an extensive game $\Gamma(\theta)$. A (pure strategy) Nash equilibrium for the game $\Gamma(\theta)$ is an element $m^* \in M$ such that, for each agent i, $g(m^*;\emptyset) \succeq_i^\theta g((m_i, m_{-i}^*);\emptyset)$ for all $m_i \in M_i$. A (pure strategy) subgame perfect equilibrium for the game $\Gamma(\theta)$ is an element $m^* \in M$ such that, for each agent i, $g(m^*;h) \succeq_i^\theta g((m_i, m_{-i}^*);h)$ for all $m_i \in M_i$ and all $h \in \mathcal{H} \backslash H_T$. Let $SPE(\Gamma(\theta))$ denote the set of subgame perfect equilibria of the game $\Gamma(\theta)$. Let also $NE(\Gamma(\theta))$ denote the set of Nash equilibria of the game $\Gamma(\theta)$.

Given a prior μ, the mechanism determines a Bayesian game $\Gamma(\mu)$ in which each player's type is his signal, and after observing his signal, player i selects a (pure) strategy from the set M_i. A strategy profile $\sigma = (\sigma_1, ..., \sigma_n)$ lists a strategy for each player i where $\sigma_i : S_i \to M_i$ and $\sigma_i(h_t, s_i)$ is the message in $M_i(h_t)$ given history h_t and signal s_i. Alternatively, we will sometimes let σ_i be a (mixed) behavior strategy i.e. a function that maps the set of possible histories and signals into the set of probability distributions over messages: $\sigma_i(\cdot \mid h_t, s_i) \in \Delta(M_i(h_t))$ is the probability distribution over $M_i(h_t)$ given history h_t and signal s_i.

An act is a mapping $\alpha : \Theta \times S \to A$. Let $\mathcal{A}$ be the set of acts. A belief is a probability β on $\Theta \times S$. In order to analyze incomplete information games, we must extend the original preferences to the ones under uncertainty. We assume that for each belief β each player i has a preference relation $\succeq_i^\beta$ over acts. We make the following assumption (which is obviously satisfied by expected utility models but much weaker than that) on this order:

Assumption 2 *Let α and $\hat{\alpha}$ be two acts, and β a belief. Then*

$$[\alpha(\theta, s) \succeq_i^\theta \hat{\alpha}(\theta, s) \text{ for all } (\theta, s) \in supp(\beta)] \Rightarrow \alpha \succeq_i^\beta \hat{\alpha},$$

where $supp(\beta)$ denotes the support of β.

Let σ be a pure strategy profile. Given a profile of pure strategies $\sigma = (\sigma_1, ..., \sigma_n)$, we will note $g(\sigma; h)$ for the act that obtains when each agent i uses strategy σ_i starting after history h occurred, i.e. each pair (θ, s) is mapped to $g(\sigma(s); h) \in A$. The act α_σ^Γ induced by σ under the mechanism Γ is defined by $\alpha_\sigma^\Gamma(\theta, s) = g(\sigma(s); \emptyset)$ for any (θ, s).

We will also assume that in the game induced by a stage mechanism, for each player best replies are always well-defined in the neighborhood of complete information when the opponents are playing according to some Nash equilibrium. In general, best-responses need not be well-defined since we allow $M_i(h)$ to be countably infinite. For instance, integer games are such an example with countably infinite message spaces in which best replies need not be well defined.[6] The next assumption ensures that in the neighborhood of complete information, against any Nash equilibrium strategy of his opponents, each player i has a strategy that is optimal at histories in some given set H and equal to some fixed strategy at every other histories.

Assumption 3 *A sequential mechanism Γ has well-defined **best replies**: for any player i, any set of histories $H \subseteq \mathcal{H}$, any $\theta \in \Theta$, any $(m_i, m_{-i}) \in NE(\Gamma(\theta))$, there exists $\bar{\xi}(i, H, \theta, m_i, m_{-i}) > 0$ such that for any $\beta \in \Delta(\Theta \times S_{-i})$ with $\beta(\theta, s_{-i}^\theta) \geq 1 - \bar{\xi}(i, H, \theta, m_i, m_{-i})$, there exists $\sigma_i^*[i, H, \theta, m_i, m_{-i}, \beta]$, or simply σ_i^*, satisfying*

$$h \notin H \Rightarrow \sigma_i^*(h; s_i^\theta) = m_i(h);$$
$$h \in H \Rightarrow g((\sigma_i^*, \sigma_{-i}); h) \succeq_i^\beta g((\sigma_i', \sigma_{-i}); h)$$

[6] If there is some player for whom there is no maximum with respect to his preference order at some state of nature, then best-replies are indeed not well-defined at this state in standard integer games.

for any $\sigma_i^{'}$ *that differs from* σ_i^* *only at* h *and any* σ_{-i} *such that* $\sigma_{-i}(s_{-i}) = m_{-i}$ *for any* s_{-i} *with* $\beta(s_{-i}) > 0$.

Remark: Note that Assumption 3 jointly restricts preference orders under uncertainty and the class of mechanisms to be considered. Provided that order of preferences are complete and transitive, Assumption 3 is vacuously satisfied in finite mechanisms, as for instance, the simple mechanism in Section 5 of Moore and Repullo (1988) that uses a finite set of messages. [7] If the mechanism is not finite but the set of outcomes is, again Assumption 3 is vacuously satisfied. Finally, we note that Assumption 3 is not needed in sequential mechanisms in which each agent moves only once. [8] Moore (1992) defines a simple sequential mechanism as a mechanism where each agent moves only once and moreover, only one agent moves at each stage. Although these simple mechanisms are considered to possess an even stronger justification for the use of subgame perfection, they are not robust to incomplete information in our sense.

When we perturb a complete information situation introducing a slight incomplete information, we must specify the equilibrium concept we use. In this paper we will focus on sequential equilibrium. Since our result provides necessary conditions, it will hold for any coarser equilibrium concept as for instance perfect Bayesian equilibrium, subgame perfect equilibrium. We now recall the definition of sequential equilibrium as defined in Kreps and Wilson (1982).

Sequential Equilibrium:

A *system of beliefs* of agent i is defined as a function $\phi_i : S_i \times \mathcal{H} \rightarrow \Delta(\Theta \times S_{-i})$. Let $\phi_i[(\theta, s_{-i}) \mid s_i, h_t]$ denote agent i's belief that the state (θ, s_i, s_{-i}) is realized when agent i's signal is s_i and the observed history is h_t. We will henceforth abuse notations and sometimes consider $\phi_i[(\theta, s_{-i}) \mid s_i, h_t]$ as an element of $\Delta(\Theta \times S)$. We also say a vector of beliefs $\phi = (\phi_1, \ldots, \phi_n)$ is Bayes consistent with a strategy profile σ if beliefs are updated from one stage to the next using Bayes' rule whenever possible (see Fudenberg and Tirole

[7]Recall that we have assumed that the set of state of nature is finite.

[8]One can directly check this in the definition of strategy σ ($\Sigma 2$) used in the proof of Theorem 1. More specifically, it can be checked there that for each player, Assumption 3 is only used at histories where this player has to choose a message and at which he has previously deviated from the equilibrium. By definition, in a simple sequential mechanism, there is no such history.

(1991) for its precise definition). An *assessment* is a pair (ϕ, σ) consisting of a profile of beliefs and a pure behavior strategy profile.

Definition 1 *A sequential equilibrium is an assessment (ϕ, σ) that satisfies condition (S) and (C):*
(S) Sequential rationality: *for all $i \in N,\ s_i \in S_i, h_t \in \mathcal{H}$:*

$$g(\sigma, h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}), h_t)$$

for each σ_i'.
(C) Consistency: *there exists a sequence of totally mixed strategy profiles $(\sigma_1^k, ..., \sigma_n^k)$ converging to $(\sigma_1, ..., \sigma_n)$ with Bayes consistent beliefs ϕ^k converging to ϕ.* [9]

For our main theorem, we need one more assumption.

Assumption 4 (One-Shot Deviation Principle) *A sequential mechanism Γ satisfies the **one-shot deviation principle** if, for every $i \in N, s_i \in S_i,\ h_t \in \mathcal{H}$ and consistent assessment (ϕ, σ): whenever*

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for every σ_i' that differs from σ_i only at h_t (local sequential rationality), it follows that

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for every σ_i' (sequential rationality).

Remark: Assumption 4 also jointly restricts the class of mechanisms considered as well as preference orders under uncertainty. Hendon, Jacobsen, and Sloth (1996) indeed show that the one-shot deviation principle holds for sequential equilibria in finite stage games as long as agents are expected utility maximizers.[10] However, this paper uses

[9]Convergence in the definition of consistency is taken uniformly over messages and histories. Given that the set of messages (and so the set of histories) can be countably infinite, two natural convergence notions can be used: *point-wise* convergence or *uniform* convergence. The set of sequential equilibria is smaller when one assumes uniform convergence. Hence, the use of uniform convergence strengthens our main result.

[10]Hendon, Jacobsen, and Sloth (1996) assume that for each i and h, $M_i(h)$ is finite. It is easy to check that their argument goes through in case $M_i(h)$ is countably infinite.

a class of preference orders under uncertainty that is weaker than the expected utility representation. In Appendix B, we provide very weak conditions on preference orders under uncertainty so that the one-shot deviation principle holds. Note also that the one-shot deviation principle trivially holds in mechanisms where each agent moves only once.

3 $\overline{SPE}$-implementation

Henceforth, we assume that A is an arbitrary topological space, and that $\mathcal{A} = A^{\Theta \times S}$ is endowed with the product topology. Given a mechanism Γ, we denote the sequential equilibrium correspondence by $\psi_\Gamma^{SE} : \mathcal{P} \to \mathcal{A}$ where each element α of $\psi_\Gamma^{SE}(\mu)$ is an act corresponding to some sequential equilibrium outcome of $\Gamma(\mu)$, which describes the alternative $\alpha(\theta, s)$ that will result for each (θ, s) (where SE stands for sequential equilibrium). Formally, $\psi_\Gamma^{SE}(\mu) \equiv \big\{\alpha \in \mathcal{A} : \alpha = \alpha_\sigma^\Gamma \text{ where } (\phi, \sigma) \text{ is a sequential equilibrium for some } \phi\big\}$. Let

$$\text{graph } \psi_\Gamma^{SE} \equiv \{(\mu, \alpha) : \alpha \in \psi_\Gamma^{SE}(\mu)\}.$$

The following notation will be convenient. If $\mathcal{B}$ is a set of acts such that for any selection f of $\mathcal{F}$, there is $\alpha \in \mathcal{B}$ for which $\alpha(\theta, s) = f(\theta)$ for any $(\theta, s) \in \text{supp}(\mu)$, then we will write $\mathcal{B} \sqsupset_\mu \mathcal{F}$. Further, if $\mathcal{B}$ is a set of acts such that $\alpha(\theta, s) \in \mathcal{F}(\theta)$ for each $\alpha \in \mathcal{B}$ and any $(\theta, s) \in \text{supp}(\mu)$, then we will write $\mathcal{B} \sqsubset_\mu \mathcal{F}$. If $\mathcal{B} \sqsubset_\mu \mathcal{F}$ and $\mathcal{B} \sqsupset_\mu \mathcal{F}$, then we write $\mathcal{B} =_\mu \mathcal{F}$.

Definition 2 *A stage mechanism Γ SE-implements an SCC $\mathcal{F} : \Theta \to A$ under μ if $\psi_\Gamma^{SE}(\mu) =_\mu \mathcal{F}$.*

When μ is a complete information prior, the above definition is equivalent to the standard definition of subgame perfect implementation. The next lemma is its formalization. We provide it with no proof.

Lemma 1 *Let μ be a complete information prior. A stage mechanism Γ SE-implements an SCC $\mathcal{F} : \Theta \to A$ under μ if and only if for each $(\theta, s^\theta) \in \Theta \times S$ with $\mu(\theta, s^\theta) > 0$, we have $g(SPE(\Gamma(\theta)); \emptyset) = \mathcal{F}(\theta)$,*

As in Chung and Ely (2003), we consider the "closure" of the solution correspondence ψ_Γ^{SE}. Define

$$\overline{\psi_\Gamma^{SE}}(\mu) = \{\alpha : (\mu, \alpha) \in \overline{\text{graph } \psi_\Gamma^{SE}}\}.$$

Recall that $(\mu, \alpha) \in \overline{\text{graph } \psi_\Gamma^{SE}}$ if there exists a sequence $\{(\mu^k, \alpha^k)\}_{k=1}^{\infty}$ such that (i) $(\mu^k, \alpha^k) \in \text{graph } \psi_\Gamma^{SE}$ for each k and (ii) $(\mu^k, \alpha^k) \to (\mu, \alpha)$. The following is our definition of robust implementation, denoted $\overline{SPE}$-*implementation*.

Definition 3 *A mechanism* Γ $\overline{SE}$*-implements an SCC* $\mathcal{F} : \Theta \to A$ *under* μ *if* $\overline{\psi_\Gamma^{SE}}(\mu) =_\mu \mathcal{F}$*. When* μ *is a complete information prior, we say that* Γ $\overline{SPE}$*-implements* $\mathcal{F}$ *under* μ*. Finally we say that an SCC* $\mathcal{F} : \Theta \to A$ *is* $\overline{SPE}$*-implementable under complete information if there exists a mechanism* Γ *that* $\overline{SE}$*-implements* $\mathcal{F}$ *under any complete information prior* μ*.*

The notion we defined above for sequential equilibria can be defined for any solution concept (as done for instance in Chung and Ely (2003)). Hence, given an arbitrary solution concept $\mathcal{E}$, we will sometimes say that an SCC is $\overline{\mathcal{E}}$-implementable.

4 Monotonicity as a Necessary Condition

In this section, we state our main theorem and illustrate the main idea of its proof via an example. We relegate the proof of the theorem to Appendix A.

4.1 Theorem and Illustration

We now recall the definition of monotonicity as defined in Maskin (1999).

Definition 4 *An SCC* $\mathcal{F}$ *is said to be monotonic if, for any* $\theta, \theta' \in \Theta$ *and any* $a \in \mathcal{F}(\theta)$*,*

$$(*) \qquad \forall i \in N, \forall b \in A, \; a \succeq_i^\theta b \Longrightarrow a \succeq_i^{\theta'} b,$$

we have $a \in \mathcal{F}(\theta')$*.*

We are now in a position to state our main theorem.

Theorem 1 *Suppose that Assumptions* 1, 2, 3 *and* 4 *are satisfied. If an SCC is* $\overline{SPE}$*-implementable under complete information, it is necessarily monotonic.*

Remark: This result seems to contradict Proposition 2 of Kreps and Wilson (1982), which shows that the sequential equilibrium correspondence is upper hemi-continuous. This apparent inconsistency comes from the very fact that the sequential equilibrium correspondence is upper hemi-continuous provided that μ has full support over $\Theta \times S$ (as is assumed in Kreps and Wilson (1982)). However – as shown in our illustration – when μ assigns probability 0 to some profile (θ, s), upper hemi-continuity may not hold.

The proof of Theorem 1 is relegated to Appendix A. Here, we rather illustrate the main idea of the proof through the simple mechanism proposed in Section 5 of Moore and Repullo (1988). The set of payoff states is $\{\theta, \theta'\}$. There are two agents, called 1 and 2. For each $i = 1, 2$, agent i's complete and transitive preference relation in state $\tilde{\theta} \in \{\theta, \theta'\}$ is given by $\succeq_i^{\tilde{\theta}}$. The agents commonly observe the state, but the planer does not observe it.

We extend the set of outcomes A to $\tilde{A} \equiv A \times \mathbb{R}^2$. An element of $\tilde{A}$ is now a tuple (a, t_1, t_2) where a is an outcome while for each player i : t_i denotes the transfer to player i. For any $\tilde{\theta} \in \{\theta, \theta'\}$, preferences over A are extended to (complete and transitive) preferences over $\tilde{A}$ denoted by $\trianglerighteq_i^{\tilde{\theta}}$. We assume that transfers to player $-i$ do not affect player i's ordering, hence, throughout this example, when considering i's evaluations over outcomes, we ignore agent $j(\neq i)$'s monetary transfer from the expression, i.e. we will abuse notations and for instance, simply note (a, t_i) instead of (a, t_i, t_j). We will further assume that for any $\tilde{\theta} \in \{\theta, \theta'\}$, $a \succ_i^{\tilde{\theta}} b$ implies $(a, t_i) \triangleright_i^{\tilde{\theta}} (b, t_i)$ for any t_i. To fix ideas, one instance of this extension is the setting with transfers and quasilinear preferences.

We assume that $f(\theta) \neq f(\theta')$ and $f : \Theta \to \tilde{A}$ is "non-monotonic" and therefore not Nash implementable. With this, we must satisfy the following condition:

$$\forall i, \forall b \in \tilde{A} : f(\theta) \trianglerighteq_i^{\theta} b \Rightarrow f(\theta) \trianglerighteq_i^{\theta'} b \quad (**)$$

Following Section 5 of Moore and Repullo (1988), we argue that this non-monotonic f can be implemented as the unique subgame perfect equilibrium outcome of the following 3-stage mechanism, under some assumptions (specified later) that are naturally satisfied in a setting with (large) transfers and quasi-linear preferences.

Stage 1: Agent 1 announces the state θ (resp., θ'). Then, the game moves to Stage 2.
Stage 2: If agent 2 agrees (i.e. announces the same state as agent 1), then the game ends here and $f(\theta)$ (resp., $f(\theta')$) is chosen. If agent 2 challenges by announcing θ' (resp., θ), the game moves to Stage 3.

Stage 3: Conditioning on agent 1's announcement θ (resp., θ') at Stage 1, agent 1 has to choose between $x(\theta)$ (resp., $x(\theta')$) and $y(\theta)$ (resp., $y(\theta')$) such that

$$\begin{gathered} x(\theta) \succ_1^{\theta} y(\theta), \text{ and} \\ (\text{resp., } x(\theta') \succ_1^{\theta'} y(\theta'), \text{ and}) \\ y(\theta) \succ_1^{\theta'} x(\theta). \\ (\text{resp., } y(\theta') \succ_1^{\theta} x(\theta').) \end{gathered}$$

Further, if agent 1 chooses $x(\theta)$ (resp., $x(\theta')$), then agent 1 receives $(x(\theta), -\Delta)$ (resp., $(x(\theta'), -\Delta)$); agent 2 receives $(x(\theta), -\Delta)$ (resp., $(x(\theta), -\Delta)$); and the planner nets 2Δ – whereas if agent 1 chooses $y(\theta)$ (resp., $y(\theta')$), then agent 1 receives $(y(\theta), -\Delta)$ (resp., $(y(\theta'), -\Delta)$); agent 2 receives $(y(\theta), +\Delta)$ (resp., $(y(\theta'), +\Delta)$); and the planner breaks even. [11] The game stops here. It is assumed that Δ is "large enough" i.e., Δ satisfies $(f(\theta'), 0) \rhd_1^{\theta'} (y(\theta), -\Delta)$; $(y(\theta), +\Delta) \rhd_2^{\theta'} (f(\theta), 0)$; and $(f(\theta'), 0) \rhd_2^{\theta'} (x(\theta'), -\Delta)$. Similarly, $(f(\theta), 0) \rhd_1^{\theta} (y(\theta'), -\Delta)$; $(y(\theta'), +\Delta) \rhd_2^{\theta} (f(\theta'), 0)$; and $(f(\theta), 0) \rhd_2^{\theta} (x(\theta), -\Delta)$. Note that by transitivity, this implies in particular that $(f(\theta'), 0) \rhd_1^{\theta'} (x(\theta), -\Delta)$ and $(f(\theta), 0) \rhd_1^{\theta} (x(\theta'), -\Delta)$.

The key property of this mechanism is that whatever the state is, there is a unique subgame perfect equilibrium where agent 1 tells the truth and agent 2 does not challenge. In addition, if agent 1 lies announcing $\tilde{\theta}$, then agent 2 challenges and at stage 3 agent 1 chooses $y(\tilde{\theta})$ while if he tells the truth, he chooses $x(\tilde{\theta})$. This can be formally described as follows.

Denote by $m_i^*(\tilde{\theta}; h)$ agent i's strategy in state $\tilde{\theta} \in \{\theta, \theta'\}$ at history h.

- $m_1^*(\theta; \emptyset) = \theta$ and $m_1^*(\theta'; \emptyset) = \theta'$;
- $m_2^*(\theta; \theta) = \theta; m_2^*(\theta'; \theta') = \theta', m_2^*(\theta; \theta') = \theta$; and $m_2^*(\theta'; \theta) = \theta'$; and
- $m_1^*(\theta; (\theta, \theta')) = x(\theta); m_1^*(\theta; (\theta', \theta)) = y(\theta'); m_1^*(\theta'; (\theta', \theta)) = x(\theta')$; and $m_1^*(\theta'; (\theta, \theta')) = y(\theta)$.

To see that m^* constitutes the unique subgame perfect equilibrium. First, note that m^* prescribes the outcome where agent 1 will announce the true state and agent 2 will

[11] The existence of such $x(\cdot)$ and $y(\cdot)$ is guaranteed by the following weak domain restriction: there are $a, b \in A$ and an agent i for whom $a \succ_i^{\theta} b$ and $b \succ_i^{\theta'} a$ (preference reversal).

not challenge. Suppose that agent 1 announces the state θ (resp., θ'). If agent 1 lies, then agent 2 can challenge her with the truth, and at stage 3, agent 1 will choose $y(\theta)$ (resp., $y(\theta')$). This is so by construction. Given the choice of Δ, this must be worse for agent 1 than what the social choice function f offers if agent 1 tells the truth. Equally, given the definition of Δ, agent 2 will be satisfied with his reward of Δ. On the other hand, if agent 1 tells the truth, then agent 2 will not (falsely) challenge, since agent 1 would now choose $x(\theta)$ (resp., $x(\theta')$) at Stage 3, which incurs a penalty of Δ for agent 2.

Hence, the above mechanism implements the SCF f in subgame perfect equilibria. By Maskin (1999), it is not implemented in Nash equilibria since f is non-monotonic. Thus, there is a Nash equilibrium outcome that does not yield the right outcome. We will show that the introduction of an arbitrarily small incomplete information (together with the failure of monotonicity) makes this "bad" Nash equilibrium outcome a sequential equilibrium outcome.

The complete information setting described above is seen as an incomplete information situation where agents have a common prior such that $\mu(\theta, s_1^\theta, s_2^\theta) = p$ and $\mu(\theta', s_1^{\theta'}, s_2^{\theta'}) = 1 - p$, where $0 < p < 1$. Now let us introduce the following perturbation of the complete information structure ν^ε.[12]

ν^ε	s_1^θ, s_2^θ	$s_1^\theta, s_2^{\theta'}$	$s_1^{\theta'}, s_2^\theta$	$s_1^{\theta'}, s_2^{\theta'}$
θ	$p(1-\varepsilon)$	$p\varepsilon/2$	$p\varepsilon/2$	0
θ'	0	0	0	$1-p$

Observe that $\nu^\varepsilon \to \mu$ as $\varepsilon \to 0$. In the following lines, we show that the failure of monotonicity allows us to turn the "bad" Nash equilibrium outcome into a "bad" sequential equilibrium outcome. We propose the following strategy profile σ^* of the game $\Gamma(\nu^\varepsilon)$ under which agent 1 claims that the true state is θ independently of his signal and player 2 never challenges player 1. More precisely, the description of the strategy is given by:

- $\sigma_1^*(s_1^\theta, \emptyset) = \sigma_1^*(s_1^{\theta'}, \emptyset) = \theta$;
- $\sigma_2^*(s_2, \tilde{\theta}) = \theta$ for any $\tilde{\theta} \in \{\theta, \theta'\}$ and any $s_2 \in \{s_2^\theta, s_2^{\theta'}\}$; and
- $\sigma_1^*(s_1^\theta, (\theta, \theta')) = x(\theta); \sigma_1^*(s_1^{\theta'}, (\theta, \theta')) = x(\theta); \sigma_1^*(s_1^\theta, (\theta', \theta)) = y(\theta')$; and $\sigma_1^*(s_1^{\theta'}, (\theta', \theta)) = x(\theta')$.

[12]The common prior assumption is completely dispensable for the rest of arguments.

Note that, $\alpha^{\Gamma}_{\sigma^*}$, the act induced by σ^*, is such that $\alpha^{\Gamma}_{\sigma^*}(\theta', s_1^{\theta'}, s_2^{\theta'}) = f(\theta)$. Hence, if each player i receives a signal $s_i^{\theta'}$ and plays according to σ_i^*, the outcome provided is $f(\theta)$.

Now we turn to the description of beliefs of players. The fundamental property we need is that when agent i has the opportunity to move after a history that is not consistent with his opponent choosing an equilibrium strategy, then this agent will assign probability one to $(\theta, s_{-i}^{\theta})$. This proposed system of beliefs will turn out to be perfectly consistent as explained further. Let us be more formal. For each player i, his belief ϕ_i^* is defined as follows:

- $\phi_i^*[\cdot|s_i, \emptyset] = \nu^{\varepsilon}(\cdot|s_i)$ for each $i = 1, 2$ and each $s_i \in \{s_i^{\theta}, s_i^{\theta'}\}$;
- $\phi_2^*[\cdot|s_2, \theta] = \nu^{\varepsilon}(\cdot|s_2)$ for each $s_2 \in \{s_2^{\theta}, s_2^{\theta'}\}$; and $\phi_2^*[\cdot|s_2, \theta'] = \delta_{(\theta, s_1^{\theta})}$ for each $s_2 \in \{s_2^{\theta}, s_2^{\theta'}\}$;
- $\phi_1^*\left[\cdot\middle|s_1, (\theta, \theta')\right] = \delta_{(\theta, s_2^{\theta})}$ for each $s_1 \in \{s_1^{\theta}, s_1^{\theta'}\}$; and $\phi_1^*[\cdot|s_1, (\theta', \theta)] = \nu^{\varepsilon}(\cdot|s_1)$ for any $s_1 \in \{s_1^{\theta}, s_1^{\theta'}\}$.

What we want to show is that the proposed assessment (ϕ^*, σ^*) constitutes a sequential equilibrium of the game $\Gamma(\nu^{\varepsilon})$ for any $\varepsilon > 0$ small enough. In this case, since $\alpha^{\Gamma}_{\sigma^*}(\theta', s_1^{\theta'}, s_2^{\theta'}) = f(\theta)$ and $\nu^{\varepsilon}(\theta', s_1^{\theta'}, s_2^{\theta'}) = 1 - p > 0$, this shows that with probability $1-p$, a bad outcome is provided (i.e. $f(\theta)$ instead of $f(\theta')$); this is indeed enough to show that the mechanism provided in this section does not $\overline{SPE}$-implements f.

First, we will check sequential rationality of (ϕ^*, σ^*). At $h_3 = (\theta, \theta')$, agent 1 has to choose between $x(\theta)$ and $y(\theta)$. Due to the construction of ϕ_1^*, regardless of the signal received, agent 1 believes with probability one that the state is θ. Then, by Assumption 2 and[13] by construction of $x(\theta)$ and $y(\theta)$, it is optimal for her to choose $x(\theta)$. Let $h_3 = (\theta', \theta)$. Suppose agent 1 received s_1^{θ}. In this case, by construction of ϕ_1^* and $\nu^{\varepsilon}(\cdot|s_1^{\theta})$, agent 1 knows that the state is θ. Here, agent 1 has to choose between $x(\theta')$ and $y(\theta')$. By construction (and Assumption 2), it is optimal for her to choose $y(\theta')$, regardless of ε. Suppose that agent 1 received $s_1^{\theta'}$. Due to the construction of ϕ_1^* and small enough $\varepsilon > 0$, agent 1 believes

[13] Assumption 2 ensures that preferences in this degenerate incomplete information case (where player i believes with probability one that the state is θ) are the same as under complete information, namely as $\trianglerighteq_i^{\theta}$.

with high probability that the state is θ'. With an additional assumption of continuity of preferences, we proceed to argue that it is optimal for her to choose $x(\theta')$. [14]

With this in mind, we move to Stage 2. Suppose that $h_2 = \theta$. In this case, if agent 2 chooses θ', he knows that agent 1 will choose $x(\theta)$. Assume that agent 2 received s_2^{θ}. In this case, by construction of ϕ_2^* and $\nu^{\varepsilon}(\cdot|s_2^{\theta})$, agent 2 knows that the state is θ. But since $(f(\theta), 0) \rhd_2^{\theta} (x(\theta), -\Delta)$, by Assumption 2, we can conclude that it is optimal for agent 2 to choose θ.

Assume that agent 2 received $s_2^{\theta'}$. As we argued before, agent 2 knows that agent 1 will choose $x(\theta)$. We also know that $(f(\theta), 0) \rhd_2^{\theta} (x(\theta), -\Delta)$. By condition $(**)$, we can obtain that $(f(\theta), 0) \unrhd_2^{\theta'} (x(\theta), -\Delta)$ as well. Since $\nu^{\varepsilon}(\cdot|s_2^{\theta'})$ assigns strictly positive weights only to $(\theta, s_1^{\theta}, s_2^{\theta'})$ and $(\theta', s_1^{\theta'}, s_2^{\theta'})$, by Assumption 2, we can conclude that it is again optimal for agent 2 to choose θ.

Suppose that $h_2 = \theta'$. In this case, due to the construction of ϕ_2^*, agent 2 believes with probability one that the state is θ and that agent 1 will choose $y(\theta')$ at Stage 3. But we know that $(y(\theta'), +\Delta) \rhd_2^{\theta} (f(\theta'), 0)$. By Assumption 2, we can conclude that for any $s_2 \in \{s_2^{\theta}, s_2^{\theta'}\}$, it is optimal for agent 2 to choose θ.

Finally, we move to Stage 1. If agent 1 chooses θ, she knows that agent 2 will choose θ so that $f(\theta)$ is chosen. On the other hand, suppose agent 1 chooses θ'. Assume also that she received s_1^{θ}. Then, she knows that the state is θ and that agent 2 will choose θ at Stage 2 and she herself will choose either $x(\theta')$ or $y(\theta')$ at Stage 3. We know that $(f(\theta), 0) \rhd_1^{\theta} (x(\theta'), -\Delta)$ and $(f(\theta), 0) \rhd_1^{\theta} (y(\theta'), -\Delta)$, hence, (by Assumption 2) we can conclude that it is optimal for her to choose θ.

Assume, on the contrary, that agent 1 received $s_1^{\theta'}$. If agent 1 deviates to θ', she knows that agent 2 will choose θ at Stage 2 and she herself will choose either $x(\theta')$ or $y(\theta')$ at Stage 3. As we argued above, we have chosen $\Delta > 0$ so that $(f(\theta), 0) \rhd_1^{\theta} (x(\theta'), -\Delta)$ and $(f(\theta), 0) \rhd_1^{\theta} (y(\theta'), -\Delta)$. By condition $(**)$, we also obtain $(f(\theta), 0) \unrhd_1^{\theta'} (x(\theta'), -\Delta)$ and $(f(\theta), 0) \unrhd_1^{\theta'} (y(\theta'), -\Delta)$. Since $\phi_1^*[\cdot|s_1^{\theta'}, \emptyset]$ assigns strictly positive weights only to $(\theta, s_1^{\theta'}, s_2^{\theta})$ and $(\theta', s_1^{\theta'}, s_2^{\theta'})$, by Assumption 2, we can conclude that it is optimal for agent 1 to choose θ at Stage 1.

[14] As shown in the proof of Theorem 1, the same argument can go through even if (perhaps due to the lack of continuity of preferences) $y(\theta')$ is a best reply. In general, what really matters here is the existence of a best response for agent 1, and this is where Assumption 3 plays a role. We checked, by means of an example, that our construction of the "bad" sequential equilibrium in the proof of Theorem 1 fails when Assumption 3 is dropped; the example is available upon request.

We conclude that (ϕ^*, σ^*) so constructed satisfies sequential rationality.

Next we will check consistency of (ϕ^*, σ^*). Roughly, given that agent 1 pools at stage 1 (i.e. agent 1 plays θ independently of his signal), if agent 2 receives an opportunity to move when agent 1 has deviated from the equilibrium path, his beliefs induced by sequences of mixed strategies converging to σ^* are going to depend on the likelihood ratio of making "mistakes" (i.e. of playing θ' for agent 1 at stage 1) for one signal over another. For instance, in case it is infinitely more likely that agent 1 makes a mistake when receiving signal s_1^{θ} rather than when she receives $s_1^{\theta'}$, then, when agent 2 has an opportunity to move after agent 1 has deviated from the equilibrium path, he will (Bayes-consistently) believe with probability one that agent 1 has received signal s_1^{θ} and so by construction that the true state must be θ. Since agent 2 also pools at stage 2, a similar argument applies for agent 1's beliefs at stage 3. Hence, many different off-the-equilibrium-path beliefs are going to satisfy consistency, and in particular the one we built above. In general, this shows that a small amount of incomplete information may induce large changes on the set of consistent beliefs.

We now move to the more technical part and construct the appropriate sequence of completely mixed strategies. Let $\{\eta_k\}_{k=1}^{\infty}$ be a sequence such that $\eta_k > 0$ for each k and $\eta_k \to 0$ as $k \to \infty$. Let a sequence of totally mixed strategy profiles $\{\sigma^k\}_{k=1}^{\infty}$ be defined as follows:

$$
\begin{aligned}
\sigma_1^k(\theta \mid \emptyset, s_1^{\theta}) &= 1 - \eta_k \\
\sigma_1^k(\theta \mid \emptyset, s_1^{\theta'}) &= 1 - \eta_k^2 \\
\sigma_2^k(\theta \mid \theta, s_2^{\theta}) = \sigma_2^k(\theta \mid \theta', s_2^{\theta}) &= 1 - \eta_k \\
\sigma_2^k(\theta \mid \theta, s_2^{\theta'}) = \sigma_2^k(\theta \mid \theta', s_2^{\theta'}) &= 1 - \eta_k^2
\end{aligned}
$$

and

$$
\begin{aligned}
\sigma_1^k(x(\theta) \mid (\theta, \theta'), s_1^{\theta}) &= 1 - \eta_k \\
\sigma_1^k(y(\theta') \mid (\theta', \theta), s_1^{\theta}) &= 1 - \eta_k \\
\sigma_1^k(x(\theta) \mid (\theta, \theta'), s_1^{\theta'}) \ (resp., \sigma_1^k(x(\theta') \mid (\theta', \theta), s_1^{\theta'}) &= 1 - \eta_k^2
\end{aligned}
$$

Note that $\sigma^k \to \sigma^*$ by construction. We can define the Bayes consistent belief profile ϕ^k associated with σ^k. We claim that $\phi^k \to \phi^*$ as $k \to \infty$. For simplicity, we only pay attention to checking beliefs of each player i after histories in which i's opponent did not play according to his equilibrium strategies. This can be done by explicitly computing the following:

$$
\begin{aligned}
& \phi_1^k[(\theta, s_2^\theta) | s_1^\theta, (\theta, \theta')] \\
= \quad & \frac{\nu^\varepsilon(\theta, s_1^\theta, s_2^\theta) \times \sigma_1^k(\theta \mid \emptyset, s_1^\theta) \times \sigma_2^k(\theta' \mid \theta, s_2^\theta)}{\nu^\varepsilon(\theta, s_1^\theta, s_2^\theta) \times \sigma_1^k(\theta \mid \emptyset, s_1^\theta) \times \sigma_2^k(\theta' \mid \theta, s_2^\theta) + \nu^\varepsilon(\theta, s_1^\theta, s_2^{\theta'}) \times \sigma_1^k(\theta \mid \emptyset, s_1^\theta) \times \sigma_2^k(\theta' \mid \theta, s_2^{\theta'})} \\
= \quad & \frac{p(1-\varepsilon)(1-\eta_k)\eta_k}{p(1-\varepsilon)(1-\eta_k)\eta_k + (p\varepsilon/2)(1-\eta_k)\eta_k^2} = \frac{p(1-\varepsilon)}{p(1-\varepsilon) + p\varepsilon\eta_k/2} \to 1 \ \ (\text{as } k \to \infty)
\end{aligned}
$$

$$
\begin{aligned}
\phi_1^k[(\theta, s_2^\theta) | s_1^{\theta'}, (\theta, \theta')] &= \frac{(p\varepsilon/2)(1-\eta_k^2)(\eta_k)}{(p\varepsilon/2)(1-\eta_k^2)(\eta_k) + (1-p)(1-\eta_k^2)\eta_k^2} = \frac{p\varepsilon/2}{p\varepsilon/2 + (1-p)\eta_k} \to 1 \ \ (\text{as } k \to \infty) \\
\phi_2^k[(\theta, s_1^\theta) | s_2^\theta, \theta'] &= \frac{p(1-\varepsilon)\eta_k}{p(1-\varepsilon)\eta_k + (p\varepsilon/2)\eta_k^2} = \frac{p(1-\varepsilon)}{p(1-\varepsilon) + p\varepsilon\eta_k/2} \to 1 \ \ (\text{as } k \to \infty) \\
\phi_2^k[(\theta, s_1^\theta) | s_2^{\theta'}, \theta'] &= \frac{(p\varepsilon/2)\eta_k}{(p\varepsilon/2)\eta_k + (1-p)\eta_k^2} = \frac{p\varepsilon/2}{p\varepsilon/2 + (1-p)\eta_k} \to 1 \ \ (\text{as } k \to \infty)
\end{aligned}
$$

5 Concluding Remarks

In this paper, we prove a necessary condition result focusing on subgame perfect implementation which is similar to the one found by Chung and Ely (2003) for undominated Nash implementation. It is natural to check what strengthening of Maskin's monotonicity would ensure $\overline{SPE}$-implementation. Given that we will have to assume monotonicity, there is probably very little gain to build a sequential mechanism, a static one would most likely be enough. Chung and Ely (2003) have provided a slight strengthening of Maskin's sufficient conditions for Nash implementability[15] under which $\overline{UNE}$-implementation is ensured (where UNE stands for undominated Nash equilibrium). Following their proof, it is very easy to check that their sufficient conditions actually imply $\overline{NE}$-implementation which in the static case is equivalent to our notion of $\overline{SPE}$-implementation.

To prove our main theorem, we restricted our attention to a class of mechanisms where best-responses are well-defined (Assumption 3). This assumption is useful in our construction and allows us to build a sequential equilibrium that does not provide the desired outcomes. There might be some possibility for non-monotonic SCCs to be $\overline{SPE}$-implementable using mechanisms where best-responses are not well-defined. While this

[15]This sufficient condition is provided under a slight strengthening of the assumptions on preference orders.

is an interesting (open) theoretical question, such a robustness result would critically use the fact that the mechanism is not well-behaved and so should not be taken too seriously.

As a final remark, we investigate the relevance of our result to the *hold-up* problem, through an example. [16] This is indeed the main theme of AFH (2009). Therefore, from the outset, we shall clarify the difference between what we do and AFH (2009). We argue that $\overline{SPE}$ implementation requires that an SCF be monotonic, and in particular, constant in the example discussed below. Then, we can conclude that the first-best investment cannot be achieved because we have to consider only constant SCFs (non-contingent contracts).[17] Furthermore, we can also show that when $\overline{SPE}$-implementation is considered, the optimal contract must be non-contingent. Arguably, non-contingent contracts are incomplete, which exhibits a stark contrast with very sophisticated contracts Maskin and Tirole (1999a) proposed. Note that Maskin and Tirole's (1999a) argument is based on the standard subgame perfect implementation.

We come to the analysis. There are two parties, a buyer (B) and seller (S) of a single unit of an indivisible good. If trade occurs, then B's payoff is $V_B = \theta - p$, where θ is the value of the good and p is the price. S's payoff is $V_S = p$. The good can be of either high or low quality. If it is high quality, then B values the good at $\theta = \theta_H = 14$, and if it is low quality, then $\theta = \theta_L = 10$.

The set of social alternatives in this economy can be defined as

$$A = \left\{(q, y_B, y_S) \in [0,1] \times \mathbb{R}^2 | \; y_B + y_S = 0\right\},$$

where q denotes the probability that the good is transferred from S to B and $y_B(y_S)$ denotes the amount of money received by $B(S)$. Agents are assumed to value (non-deterministic) outcomes through expected utility.

First, we claim that only constant SCFs are monotonic. Let an arbitrary SCF f^* be

[16] The example is adapted from AFH (2007) who acknowledge that their example is based on the one of Hart and Moore (2003).

[17] Following Fudenberg, Kreps, and Levine (1988), AFH (2009) argue that the hold-up problem disappears for some crazy type perturbations, which is much richer perturbations than this paper admits (See also a brief discussion on this point in our introduction). On the other hand, they argue that the hold-up problem remains for some common p-belief perturbations, which we use as the same class of perturbations. The reader should be referred to AFH (2009) for the details.

that $f^*(\theta_L) = (q^*, y_B^*, y_S^*) \in A$. Let $\theta = \theta_L$ and $\theta' = \theta_H$. Let $b = (q, y_B, y_S) \in A$.

$$
\begin{aligned}
& f^*(\theta_L) \succeq_B^{\theta_L} b \text{ and } f^*(\theta_L) \succeq_S^{\theta_L} b \\
\Longrightarrow\; & 10q^* + y_B^* \geq 10q + y_B \text{ and } y_S^* \geq y_S \\
\Longrightarrow\; & y_B^* \geq 10(q - q^*) + y_B \text{ and } y_S^* \geq y_S \\
\Longrightarrow\; & -y_S^* \geq 10(q - q^*) + y_B \text{ and } y_S^* \geq y_S \text{ because } y_B^* + y_S^* = 0 \\
\Longrightarrow\; & -y_S \geq 10(q - q^*) + y_B \text{ and } y_S^* \geq y_S \text{ because } y_S^* \geq y_S \text{ iff } -y_S \geq -y_S^* \\
\Longrightarrow\; & 0 \geq 10(q - q^*) \text{ because } y_B + y_S = 0 \\
\Longrightarrow\; & q^* \geq q \text{ and } 10q^* + y_B^* \geq 10q + y_B \\
\Longrightarrow\; & q^* \geq q \text{ and } 14q^* + y_B^* \geq 14q + y_B \\
\Longrightarrow\; & f^*(\theta_L) \succeq_B^{\theta_H} b \text{ and } f^*(\theta_L) \succeq_S^{\theta_H} b
\end{aligned}
$$

This implies that $(*)$ in the definition of monotonicity holds between θ_L and θ_H. Then, monotonicity requires that $f^*(\theta_L) = f^*(\theta_H)$.

Second, we embed this example into the hold-up problem. Here, any implementing mechanism is interpreted as a *contract*. In this example, any monotonic SCF corresponds to a non-contingent contract. There are four dates in this contractual relationship: At date 1, the two parties sign a contract, i.e., propose a mechanism and agree on playing it later. At date 2, S makes a "non-contractible" (relation-specific) investment. At date 3, the parties receive the signals. At date 4, the parties play the mechanism which is proposed at date 1. We shall enlarge on the investment stage of the model. After a contract is signed, only S makes a non-contractible investment e_S, which increases the probability that the good entails high value.[18] For simplicity, there are only two levels of investment: either $e_S = 1$ (investment) or $e_S = 0$ (no investment). The cost of investment $c(e_S)$ is given as: $c(e_S) = 1/3$ if $e_S = 1$ and $c(e_S) = 0$ if $e_S = 0$. It is common knowledge that the likelihood

[18]Here the seller's investment entails direct externalities à la Che and Hausch (1999). This externality is important for the rest of the argument because if the buyer instead makes investment (i.e., selfish investment), there is no underinvestment in the hold-up problem, given the monotonic (and so constant) SCF. Clearly, this is due to the property that, in the present setting, a monotonic SCF is constant; whether this is a general property in hold-up problems is left as an open question.

of the states of the world depend upon signal realizations at date 3 and S's investment:

$$\begin{aligned}
\mu(\theta_L, s_B^{\theta_L}, s_S^{\theta_L}|e_S = 0) = \mu(\theta_H, s_B^{\theta_H}, s_S^{\theta_H}|e_S = 0) &= 1/2 \\
\mu(\theta_L, s_B^{\theta_L}, s_S^{\theta_L}|e_S = 1) &= 1/3 \\
\mu(\theta_H, s_B^{\theta_H}, s_S^{\theta_H}|e_S = 1) &= 2/3
\end{aligned}$$

Let a monotonic SCF f^{**} be such that $f^{**}(\theta_L) = f^{**}(\theta_H) = (1, -10, 10)$. Note that we already have argued that only constant SCFs are monotonic. This f^{**} is S's best contract that satisfies both ex post individual rationality and monotonicity.[19] Our Theorem 1 shows that this SCF f^{**} is S's best contract among all ex post individually rational SCFs that are $\overline{SPE}$-implementable.[20] It is rather trivial to realize that the SCF f^{**} is easily implementable in any equilibrium because it is a constant SCF and there is no need to extract information from the agents. Given the SCF f^{**}, it is optimal for S to make "no" investment. Thus, $\overline{SPE}$-implementability makes the hold-up problem to involve *underinvestment*. Che and Hausch (1999) obtain the same conclusion (their Proposition 2) in a more general setup. They assume that the parties cannot credibly commit not to renegotiate the initial non-contingent contract. These contract terms can be renegotiated to ex post efficient quantity after the parties realize the state. Restricting attention to the initial non-contingent contracts with ex post renegotiation, they show that when the degree of cooperative nature of investments is sufficiently high, any non-contingent contract has no value, i.e., the same as no contract. Indeed, within this example, we show that there is no loss of generality to restrict attention to non-contingent contracts if we require $\overline{SPE}$-implementation for the initial contract. This can be a support to the property rights approach of Grossman and Hart (1986) and Hart and Moore (1990).[21]

[19] Ex post individual rationality guarantees that each player receives at least the utility of no trade at each state of nature.

[20] To be exact, this is the case, provided that the implementing mechanism satisfies Assumption 3.

[21] One way to get positive results would be to weaken the requirement of "exact" implementation to "virtual" implementation where non-deterministic mechanisms are used and we only require the SCR to be implemented with high probability. By Abreu and Matsushima (1992), we know that we can virtually implement almost any SCF using rationalizability as the solution concept. One can indeed show that if a SCF is (virtually) R-implementable (i.e. using rationalizability as the solution concept) then, it is also (virtually) $\bar{R}$-implementable (see Kunimoto (2009) for details). While definitely of great theoretical importance, on a practical viewpoint, the Abreu and Matsushima's approach has several drawbacks. First, mechanisms used there are demanding from the agents' viewpoint. Indeed, Glazer and Rosenthal (1992) argue that these mechanisms will not perform as predicted because it may involve many rounds of iterated

Appendix

A Proof of Theorem 1

Let μ be a complete information prior such that $\mu(\tilde{\theta}, s^{\tilde{\theta}}) > 0$ for all $\tilde{\theta} \in \Theta$, and let $\mathcal{F}$ be a $\overline{SPE}$-implementable SCC with implementing mechanism Γ. Fix any $\theta, \theta^{'} \in \Theta$ and any $a \in \mathcal{F}(\theta)$. Suppose θ and θ' are two possible states satisfying $(*)$ in Definition 4 (p.11). We will show that $a \in \mathcal{F}(\theta')$.

Since Γ $\overline{SPE}$-implements $\mathcal{F}$, it must also SPE-implements $\mathcal{F}$. Thus, by Lemma 1, there exists a subgame perfect equilibrium m^*_θ in $\Gamma(\theta)$ such that $g(m^*_\theta) = a$. Clearly, m^*_θ is actually a Nash equilibrium of $\Gamma(\theta)$. From $(*)$, it follows that m^*_θ is also a Nash equilibrium of $\Gamma(\theta')$. Recall that $\mathcal{H}$ denotes the set of all possible histories. For each $t \geq 0$, let h^*_t be the history induced by m^*_θ up to date t and denote $\mathcal{H}^*$ for the set of all such histories. In addition, for each player i, let $\mathcal{H}^*_{-i}$ be the set of histories h along which every player $j \neq i$ has chosen the message $m^*_{\theta,j}(h')$; formally, $\mathcal{H}^*_{-i} \equiv \{h \in \mathcal{H} : h = (\emptyset, m^1, m^2, ..., m^{t-1})$ for some t and $m^{t'}_j = m^{*,t'}_{j,\theta}$ for all $t' \leq t-1$ and all $j \neq i\}$. Note that $h^*_t \in \mathcal{H}^*_{-i}$ for each $t \geq 1$.

Consider the following family of information structure ν^ε. For each player i, let τ_i represent the profile of signals $s = (s_1, ..., s_n)$ defined by $s_i = s^{\theta'}_i$ and $s_j = s^\theta_j$ for all $j \neq i$. For all i, ν^ε describes

$$\begin{aligned}
\nu^\varepsilon(\theta, \tau_i) &= \frac{\varepsilon}{n}\mu(\theta, s^\theta); \\
\nu^\varepsilon(\theta, s^\theta) &= (1-\varepsilon)\mu(\theta, s^\theta); \text{ and} \\
\nu^\varepsilon(\tilde{\theta}, s^{\tilde{\theta}}) &= \mu(\tilde{\theta}, s^{\tilde{\theta}}) \ \ \forall \tilde{\theta} \neq \theta.
\end{aligned}$$

In this information structure when the state is anything other than θ or θ', the state is common knowledge. Furthermore, when a player observes θ, he knows that the state is θ. Obviously, $\nu^\varepsilon \to \mu$ as $\varepsilon \to 0$. [22] The support of ν^ε is denoted

$$\text{supp}(\nu^\varepsilon) = \{(\tilde{\theta}, s^{\tilde{\theta}}) : \tilde{\theta} \in \Theta\} \cup \{(\theta, \tau_i) : i \in N\}.$$

dominance, suspecting that players may sometimes "abandon the logic of iterated dominance in favor of a focal point in the game". This intuition is indeed supported by experimental evidence; see Sefton and Yavas (1996). Second, and more importantly, these mechanisms may provide ex post an outcome arbitrarily inefficient, unfair, or, more generally, far from the desired alternative. For these reasons, these mechanisms have not been considered of great practical importance and, a fortiori, have not been considered as putting into question the foundation of incomplete contracts.

[22] We use exactly the same information structures as in Chung and Ely (2003).

Fix $\varepsilon > 0$ to be sufficiently small so that $\nu^{\varepsilon}((\theta', s^{\theta'}_{-i}) \mid s^{\theta'}_{i}) \geq 1 - \bar{\xi}(i, \mathcal{H}^*_{-i} \backslash \mathcal{H}^*, \theta', m^*_{i,\theta}, m^*_{-i,\theta})$ where $\bar{\xi}(i, \mathcal{H}^*_{-i} \backslash \mathcal{H}^*, \theta', m^*_{i,\theta} m^*_{-i,\theta})$ is in Assumption 3.

We build a sequential equilibrium (ϕ, σ) of $\Gamma(\nu^{\varepsilon})$ where σ induces an act α^{Γ}_{σ} for which $\alpha^{\Gamma}_{\sigma}(\theta', s^{\theta'}) = a$. Hence, this will show that $(\nu^{\varepsilon}, \alpha^{\Gamma}_{\sigma}) \in$ graph ψ^{SE}_{Γ} for all $\varepsilon > 0$ small enough. Although σ may depend on ε, we will see that the induced act α^{Γ}_{σ} does not. Hence, $(\nu^{\varepsilon}, \alpha^{\Gamma}_{\sigma}) \to (\mu, \alpha^{\Gamma}_{\sigma}) \in \overline{\text{graph } \psi^{SE}_{\Gamma}}$ as $\varepsilon \to 0$. Thus since, by our hypothesis, Γ $\overline{\psi^{SE}_{\Gamma}}$-implements $\mathcal{F}$ under μ and $\mu(\theta', s^{\theta'}) > 0$, we must have $a = \alpha^{\Gamma}_{\sigma}(\theta', s^{\theta'}) \in \mathcal{F}(\theta')$, which will complete the proof.

In the following lines, we define a strategy σ and a family of system of beliefs Φ so that σ induces an act α^{Γ}_{σ} for which $\alpha^{\Gamma}_{\sigma}(\theta', s^{\theta'}) = a$. In addition, we will show that (ϕ, σ) is a sequential equilibrium of $\Gamma(\nu^{\varepsilon})$ for some $\phi \in \Phi$. Φ and σ are defined as follows:

Definition of Φ:

$\phi \in \Phi$ if and only ϕ satisfies the following three properties.

Φ1. Fix any $i \in N$, any $h_t \notin \mathcal{H}^*_{-i}$,

$$\phi_i \left[\cdot | s^{\theta'}_i, h_t \right] = \delta_{(\theta, s^{\theta}_{-i})}$$

also

$$\text{supp} \left(\phi_i \left[\cdot | s^{\theta}_i, h_t \right] \right) \subseteq \text{supp} \left(\nu^{\varepsilon} \left[\cdot | s^{\theta}_i \right] \right)$$

and for all $l \neq i$ with $h_t \in \mathcal{H}^*_{-l}$: (i.e., l has deviated)

$$\phi_i[(\theta, \tau_l) \mid s^{\theta}_i, h_t] = 0.$$

Φ2. For any $i \in N$, any $h_t \in \mathcal{H}^*_{-i}$, any $s_i \in \{s^{\theta}_i, s^{\theta'}_i\}$:

$$\phi_i[\cdot | s_i, h_t] = \nu^{\varepsilon}(\cdot | s_i).$$

Φ3. For any $i \in N$, any $h_t \in \mathcal{H}$ and any $s^{\tilde{\theta}}_i \notin \{s^{\theta}_i, s^{\theta'}_i\}$, we just assume that $\phi_i \left[\cdot \mid s^{\tilde{\theta}}_i, h_t \right] = \delta_{(\tilde{\theta}, s^{\tilde{\theta}}_{-i})}$ where δ_x denotes the probability measure that puts probability 1 on $\{x\}$.

Definition of σ:

Σ1. For any player i and any $h_t \in \mathcal{H}^*$ or $h_t \notin \mathcal{H}^*_{-i}$: $\sigma_i(h_t, s^{\theta'}_i) = m^*_{i,\theta}(h_t)$;

Σ2. For any player i and any $h_t \in \mathcal{H}^*_{-i} \backslash \mathcal{H}^*$, $\sigma_i(h_t, s_i^{\theta'}) = \sigma_i^*(h_t, s_i^{\theta'})$ where $\sigma_i^* = \sigma_i^*[i, \mathcal{H}^*_{-i} \backslash \mathcal{H}^*, \theta', m^*_{i,\theta}, m^*_{-i,\theta}, \nu^\varepsilon[\cdot \mid s_i^{\theta'}]]$ as defined in Assumption 3 and so satisfies:

$$h \in \mathcal{H}^* \text{ or } h \notin \mathcal{H}^*_{-i} \Rightarrow \sigma_i^*(h, s_i^{\theta'}) = m^*_{i,\theta}(h);$$

$$h \in \mathcal{H}^*_{-i} \backslash \mathcal{H}^* \Rightarrow g((\sigma_i^*, \hat{\sigma}_{-i}); h) \succeq_i^{\nu^\varepsilon(\cdot|s_i^{\theta'})} g((\sigma_i', \hat{\sigma}_{-i}); h)$$

for any σ_i' that differs from σ_i^* only at h (one-shot deviation) and any $\hat{\sigma}_{-i}$ satisfying $\hat{\sigma}_{-i}(s_{-i}) = m^*_{-i,\theta}$ for any s_{-i} with $\nu^\varepsilon(s_{-i}|s_i^{\theta'}) > 0$. This is well-defined by Assumption 3 because ε is small enough so that $\nu^\varepsilon(\theta', s_{-i}^{\theta'}|s_i^{\theta'}) \geq 1 - \bar{\xi}(i, \mathcal{H}^*_{-i} \backslash \mathcal{H}^*, \theta', m^*_{i,\theta}, m^*_{-i,\theta}, \nu^\varepsilon[\cdot \mid s_i^{\theta'}]);$

Σ3. For any player i and any $h_t \in \mathcal{H} : \sigma_i(h_t, s_i^\theta) = m^*_{i,\theta}(h_t);$

Σ4. And for any $h_t \in \mathcal{H}$, $\sigma_i(h_t, s_i^{\tilde{\theta}}) = m^*_{\tilde{\theta},i}(h_t)$ for $\tilde{\theta} \neq \theta, \theta'$ where $m^*_{\tilde{\theta}}$ is an arbitrary subgame perfect equilibrium of $\Gamma(\tilde{\theta})$. This is well-defined since $\mathcal{F}$ is implementable in subgame perfect equilibrium under complete information.

Note that $h_T[\sigma(s^{\theta'}), \emptyset] = h_T[m^*_\theta, \emptyset]$ and so, σ generates an act α_σ^Γ (that does not depend on ε) for which $\alpha_\sigma^\Gamma(\theta', s^{\theta'}) = g(\sigma(s^{\theta'}); \emptyset) = g(m^*_\theta; \emptyset) = a$. Hence, it only remains to show that (ϕ, σ) constitutes a sequential equilibrium for some $\phi \in \Phi$. In Section A.1, we will show that (ϕ, σ) satisfies sequential rationality for any $\phi \in \Phi$; and we will also establish that (ϕ, σ) satisfies consistency for some $\phi \in \Phi$ in Section A.2.

A.1 Sequential rationality

Fix any $\phi \in \Phi$. Sequential rationality of (ϕ, σ) will be proved by Claims 1 and 2 below.

Claim 1 *For any* $i \in N$, $s_i \neq s_i^{\theta'}, h_t \in \mathcal{H}$:

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i, h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for each σ_i'.

Proof of Claim 1: Fix any player i. It is obvious for $s_i^{\tilde{\theta}} \neq s_i^\theta$ because by **Φ3**, $\phi_i\left[\cdot \mid s_i^{\tilde{\theta}}, h_t\right] = \delta_{(\tilde{\theta}, s_{-i}^{\tilde{\theta}})}$ and so state $\tilde{\theta}$ is common knowledge. By **Σ4**, we can further conclude that $\sigma(s^{\tilde{\theta}}) = m^*_{\tilde{\theta}}$ is a subgame perfect equilibrium in the complete information game $\Gamma(\tilde{\theta})$. Hence, we focus on the case where $s_i = s_i^\theta$. By construction, $\nu^\varepsilon(\theta \mid s_i^\theta) = 1$

and so this player knows that his preference is given by $\succeq_i^\theta$. The uncertainty he faces is rather on the signals of his opponents, i.e. whether the profile of signals is s^θ or τ_k for some $k \neq i$.

Let $h_t \notin \mathcal{H}^*_{-i}$. By **Σ3** we know that $\sigma(s^\theta) = m^*_\theta$. Hence, $h_T[\sigma(s^\theta), h_t] = h_T[m^*_\theta, h_t]$ and so

$$g(\sigma(s^\theta); h_t) = g(m^*_\theta; h_t).$$

In addition, for each $l \neq i$ with $h_t \notin \mathcal{H}^*_{-l}$, by **Σ1** and **Σ3** we know that $\sigma_{-i}(h_t, \tau_l) = m^*_{-i,\theta}(h_t)$. [23] For any history $h_{t'}$ that follows h_t, we must have $h_{t'} \notin \mathcal{H}^*_{-l}$. By applying again **Σ1** and **Σ3** we get that $\sigma_{-i}(\tau_l) \mid_{h_t} = m^*_{-i,\theta} \mid_{h_t}$. Hence, we obtain $h_T[\sigma(\tau_l), h_t] = h_T[m^*_\theta, h_t]$ and so for each $l \neq i$ with $h_t \notin \mathcal{H}^*_{-l}$, we have

$$g(\sigma(\tau_l); h_t) = g(m^*_\theta; h_t).$$

In case player i deviates to σ'_i, he can induce the following terminal histories: $h_T[\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}), h_t] = h_T[m'_i, m^*_{-i,\theta}, h_t]$ for some strategy m'_i and so

$$g(\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}); h_t) = g(m'_i, m^*_{-i,\theta}; h_t).$$

In addition, for each $l \neq i$ with $h_t \notin \mathcal{H}^*_{-l}$, we know that $\sigma_{-i}(\tau_l) \mid_{h_t} = m^*_{-i,\theta} \mid_{h_t}$. Hence, $h_T[\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_l), h_t] = h_T[m'_i, m^*_{-i,\theta}, h_t]$ and so for each $l \neq i$ with $h_t \notin \mathcal{H}^*_{-l}$, we have

$$g(\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_l); h_t) = g(m'_i, m^*_{-i,\theta}; h_t).$$

Since m^*_θ is a subgame perfect equilibrium in the complete information game $\Gamma(\theta)$, we have $g(m^*_\theta; h_t) \succeq_i^\theta g(m'_i, m^*_{-i,\theta}; h_t)$. Thus, we get $g(\sigma(s^\theta); h_t) \succeq_i^\theta g(\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}); h_t)$ and for each $l \neq i$ such that $h_t \notin \mathcal{H}^*_{-l} : g(\sigma(\tau_l); h_t) \succeq_i^\theta g(\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_l); h_t)$. Because by **Φ1**, $\phi_i[\cdot \mid s^\theta_i, h_t]$ may assign strictly positive weight only to (θ, s^θ_{-i}) and (θ, τ_l) for each $l \neq i$ such that $h_t \notin \mathcal{H}^*_{-l}$, we can conclude with Assumption 2

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s^\theta_i, h_t]} g((\sigma'_i, \sigma_{-i}); h_t).$$

Let $h_t \in \mathcal{H}^*_{-i}$. Let us distinguish two cases. First, assume that $h_t \in \mathcal{H}^*_{-i} \backslash \mathcal{H}^*$. Since $h_t \in \mathcal{H}^*_{-i}$ and $h_t \notin \mathcal{H}^*$, there must exist $t' < t$ such that $\sigma_i(h_{t'}, s^\theta_i) \neq m^*_{i,\theta}(h_{t'})$ where

[23] We abuse the notation because we should use $\sigma_{-i}(\tau_l \backslash s^\theta_i, h_t)$ instead of $\sigma_{-i}(\tau_l, h_t)$. This abuse will be used at several places in the paper.

$h_{t'}$ is a truncation of history h_t. Then, for any history $h_{t''}$ following $h_{t'}$ (and so in particular, following h_t), we have $h_{t''} \notin \mathcal{H}^*_{-k}$ for each $k \neq i$. By **Σ1** and **Σ3**, we thus obtain $\sigma(h_{t''}, s^\theta) = \sigma(h_{t''}, \tau_k) = m^*_\theta(h_{t''})$ for each $k \neq i$. Hence, for each $k \neq i$ we have $h_T[\sigma(s^\theta), h_t] = h_T[\sigma(\tau_k), h_t] = h_T[m^*_\theta, h_t]$, which further implies

$$g(\sigma(s^\theta); h_t) = g(\sigma(\tau_k); h_t) = g(m^*_\theta; h_t).$$

Consider the case where player i deviates to σ'_i. Here, **Σ1** and **Σ3** allow us to conclude that for each $k \neq i$, player i can induce the following terminal histories: $h_T[\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}), h_t] = h_T[\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_k), h_t] = h_T[m'_i, m^*_{-i,\theta}, h_t]$ for some strategy m'_i, which implies

$$g(\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}); h_t) = g(\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_k); h_t) = g(m'_i, m^*_{-i,\theta}; h_t).$$

Since m^*_θ is a subgame perfect equilibrium in the complete information game $\Gamma(\theta)$, we already have $g(m^*_\theta; h_t) \succeq^\theta_i g(m'_i, m^*_{-i,\theta}; h_t)$. Thus, we also get $g(\sigma(s^\theta); h_t) \succeq^\theta_i g(\sigma'_i(s^\theta_i), \sigma_{-i}(s^\theta_{-i}); h_t)$ and $g(\sigma(\tau_k); h_t) \succeq^\theta_i g(\sigma'_i(s^\theta_i), \sigma_{-i}(\tau_k); h_t)$ for each $k \neq i$. Now, since by **Φ2** we know that $\phi_i[\cdot \mid s^\theta_i, h_t]$ assigns a strictly positive weight only to (θ, s^θ_{-i}) and (θ, τ_k) for each $k \neq i$, we can conclude with Assumption 2

$$g(\sigma, h_t) \succeq_i^{\phi_i[\cdot|s^\theta_i;h_t]} g((\sigma'_i, \sigma_{-i}); h_t).$$

Consider now the second case where $h_t \in \mathcal{H}^*$. Note that $h_{t+1} = (h_t, \sigma(h_t, s^\theta)) = (h_t, \sigma(h_t, \tau_k)) = (h_t, m^*_\theta(h_t)) = h^*_{t+1} \in \mathcal{H}^*$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_t \in \mathcal{H}^*$. Similar argument can be made inductively so that any subsequent history also falls into $\mathcal{H}^*$. Thus, $h_T[\sigma(s^\theta), h_t] = h_T[\sigma(\tau_k), h_t] = h_T[m^*_\theta, h_t]$, and so we obtain

$$g(\sigma(s^\theta); h_t) = g(\sigma(\tau_k); h_t) = g(m^*_\theta; h_t).$$

Now consider that player i deviates to σ'_i. Let $\hat{t} \geq t$ be the first date at which $\sigma'_i(h_{\hat{t}}, s^\theta_i) \neq \sigma_i(h_{\hat{t}}, s^\theta_i)$; or equivalently, $\sigma'_i(h_{\hat{t}}, s^\theta_i) \neq m^*_{i,\theta}(h_{\hat{t}})$. As above, one can inductively show that as long as $t' < \hat{t}$, we obtain $h_{t'+1} = (h_{t'}, \sigma'_i(h_{t'}, s^\theta_i), \sigma_{-i}(h_{t'}, s^\theta_{-i})) = (h_{t'}, \sigma'_i(h_{t'}, s^\theta_i), \sigma_{-i}(h_{t'}, \tau_k)) = (h_{t'}, m^*_{i,\theta}(h_{t'}), m^*_{-i,\theta}(h_{t'})) \in \mathcal{H}^*$ for each $k \neq i$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_{t'} \in \mathcal{H}^*$. In addition, $h_{t'+1} \notin \mathcal{H}^*_{-k}$ for each $k \neq i$ and $t' \geq \hat{t}$. Hence, for $t' \geq \hat{t}$, $h_{t'+1} = (h_{t'}, \sigma'_i(h_{t'}, s^\theta_i), \sigma_{-i}(h_{t'}, s^\theta_{-i})) = (h_{t'}, \sigma'_i(h_{t'}, s^\theta_i), \sigma_{-i}(h_{t'}, \tau_k)) = (h_{t'}, \sigma'_i(h_{t'}, s^\theta_i), m^*_{-i,\theta}(h_{t'}))$ for each $k \neq i$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_{t'} \notin \mathcal{H}^*_{-k}$ for each $k \neq i$. So we get $h_T[\sigma_i'(s_i^\theta), \sigma_{-i}(s_{-i}^\theta), h_t] = h_T[\sigma_i'(s_i^\theta), \sigma_{-i}(\tau_k), h_t] = h_T[m_i', m_{-i,\theta}^*, h_t]$ for some strategy m_i', which implies

$$g(\sigma_i'(s_i^\theta), \sigma_{-i}(s_{-i}^\theta); h_t) = g(\sigma_i'(s_i^\theta), \sigma_{-i}(\tau_k); h_t) = g(m_i', m_{-i,\theta}^*; h_t).$$

Here again, since m_θ^* is a subgame perfect equilibrium in the complete information game $\Gamma(\theta)$, we have $g(m_\theta^*; h_t) \succeq_i^\theta g(m_i', m_{-i,\theta}^*; h_t)$. Thus, we get $g(\sigma(s^\theta); h_t) \succeq_i^\theta g(\sigma_i'(s_i^\theta), \sigma_{-i}(s_{-i}^\theta); h_t)$ and $g(\sigma(\tau_k); h_t) \succeq_i^\theta g(\sigma_i'(s_i^\theta), \sigma_{-i}(\tau_k); h_t)$ for each $k \neq i$. Now since by **Φ2**, $\phi_i[\cdot \mid s_i^\theta, h_t]$ may assign strictly positive weight only to (θ, s_{-i}^θ) and (θ, τ_k) for each $k \neq i$, we can conclude with Assumption 2

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i^\theta, h_t]} g((\sigma_i', \sigma_{-i}); h_t).$$

This completes the proof. ■

Claim 2 *For any $i \in N$, $s_i = s_i^{\theta'}$, and $h_t \in \mathcal{H}$:*

$$g(\sigma, h_t) \succeq_i^{\phi_i[\cdot|s_i, h_t]} g((\sigma_i', \sigma_{-i}), h_t)$$

for each σ_i'.

Proof of Claim 2: This claim will be proved by studying three different cases depending on the type of history we consider: (1) $h_t \notin \mathcal{H}_{-i}^*$; (2) $h_t \in \mathcal{H}^*$; and (3) $h_t \in \mathcal{H}_{-i}^* \backslash \mathcal{H}^*$.

Let us first consider the case (1) $h_t \notin \mathcal{H}_{-i}^*$. By **Σ3** we know that $\sigma_{-i}(s_{-i}^\theta) = m_{-i,\theta}^*$. In addition, for any history $h_{t'}$ following h_t, we have $h_{t'} \notin \mathcal{H}_{-i}^*$. Thus, by **Σ1**, we obtain $\sigma_i(h_{t'}, s_i^{\theta'}) = m_{i,\theta}^*(h_{t'})$ for any subsequent history $h_{t'}$. This further implies that $h_T[\sigma(s_i^{\theta'}, s_{-i}^\theta), h_t] = h_T[m_\theta^*, h_t]$ and so we obtain

$$g(\sigma(s_i^{\theta'}, s_{-i}^\theta); h_t) = g(m_\theta^*; h_t).$$

Consider that player i deviates to σ_i'. Then, we have $h_T[\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^\theta), h_t] = h_T[m_i', m_{-i,\theta}^*, h_t]$ for some strategy m_i'. Hence, we obtain

$$g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^\theta); h_t) = g(m_i', m_{-i,\theta}^*; h_t).$$

Since m_θ^* is a subgame perfect equilibrium in the complete information game $\Gamma(\theta)$, we have $g(m_\theta^*; h_t) \succeq_i^\theta g(m_i', m_{-i,\theta}^*; h_t)$. Thus, we also get $g(\sigma(s_i^{\theta'}, s_{-i}^\theta); h_t) \succeq_i^\theta g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^\theta); h_t)$. Because by **Φ1**, $\phi_i[(\theta, s_{-i}^\theta) \mid s_i^{\theta'}, h_t] = 1$, we can conclude with Assumption 2

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i^{\theta'}, h_t]} g((\sigma_i', \sigma_{-i}); h_t).$$

Consider now the case (2) $h_t \in \mathcal{H}^*$. Note that $h_{t+1} = (h_t, \sigma(h_t, s_i^{\theta'}, s_{-i}^{\theta'})) = (h_t, \sigma(h_t, s_i^{\theta'}, s_{-i}^{\theta})) = (h_t, m_\theta^*(h_t)) = h_{t+1}^* \in \mathcal{H}^*$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_t \in \mathcal{H}^*$. Similar argument can be made inductively so that any subsequent history also falls into $\mathcal{H}^*$. Hence we have $h_T[\sigma(s_i^{\theta'}, s_{-i}^{\theta'}), h_t] = h_T[\sigma(s_i^{\theta'}, s_{-i}^{\theta}), h_t] = h_T[m_\theta^*, h_t]$, which implies

$$g(\sigma(s_i^{\theta'}, s_{-i}^{\theta'}); h_t) = g(\sigma(s_i^{\theta'}, s_{-i}^{\theta}); h_t) = g(m_\theta^*; h_t).$$

Now consider that player i deviates to σ_i'. Let $\hat{t} \geq t$ be the first date at which $\sigma_i'(h_{\hat{t}}, s_i^{\theta'}) \neq \sigma_i(h_{\hat{t}}, s_i^{\theta'})$; or equivalently, $\sigma_i'(h_{\hat{t}}, s_i^{\theta'}) \neq m_{i,\theta}^*(h_{\hat{t}})$. As above, similar argument would show that as long as $t' < \hat{t}$, we have $h_{t'+1} = (h_{t'}, \sigma_i'(h_{t'}, s_i^{\theta'}), \sigma_{-i}(h_{t'}, s_{-i}^{\theta'})) = (h_{t'}, \sigma_i'(h_{t'}, s_i^{\theta'}), \sigma_{-i}(h_{t'}, s_{-i}^{\theta})) = (h_{t'}, m_{i,\theta}^*(h_{t'}), m_{-i,\theta}^*(h_{t'})) \in \mathcal{H}^*$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_{t'} \in \mathcal{H}^*$. In addition, $h_{\hat{t}+1} = (h_{\hat{t}}, \sigma_i'(h_{\hat{t}}, s_i^{\theta'}), \sigma_{-i}(h_{\hat{t}}, s_{-i}^{\theta'})) = (h_{\hat{t}}, \sigma_i'(h_{\hat{t}}, s_i^{\theta'}), \sigma_{-i}(h_{\hat{t}}, s_{-i}^{\theta})) = (h_{\hat{t}}, \sigma_i'(h_{\hat{t}}, s_i^{\theta'}), m_{-i,\theta}^*(h_{\hat{t}}))$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_{\hat{t}} \in \mathcal{H}^*$. Note that $h_{t'} \notin \mathcal{H}_{-k}^*$ for each $k \neq i$ and for $t' \geq \hat{t} + 1$. Therefore, using an inductive argument, one can show that $h_{t'+1} = (h_{t'}, \sigma_i'(h_{t'}, s_i^{\theta'}), \sigma_{-i}(h_{t'}, s_{-i}^{\theta'})) = (h_{t'}, \sigma_i'(h_{t'}, s_i^{\theta'}), \sigma_{-i}(h_{t'}, s_{-i}^{\theta})) = (h_{t'}, \sigma_i'(h_{t'}, s_i^{\theta'}), m_{-i,\theta}^*(h_{t'}))$ where the second and third equalities are assured by **Σ1** and **Σ3** and we use the fact that $h_{t'} \notin \mathcal{H}_{-k}^*$ for each $k \neq i$. So we get $h_T[\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta}), h_t] = h_T(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta'}), h_t] = h_T[m_i', m_{-i,\theta}^*, h_t]$ for some strategy m_i', which implies

$$g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta}); h_t) = g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta'}); h_t) = g(m_i', m_{-i,\theta}^*; h_t). \tag{1}$$

Here again, since m_θ^* is a subgame perfect equilibrium in the complete information game $\Gamma(\theta)$, we have $g(m_\theta^*; h_t) \succeq_i^\theta g(m_i', m_{-i,\theta}^*; h_t)$. Thus, we also get

$$g(\sigma(s_i^{\theta'}, s_{-i}^{\theta}); h_t) \succeq_i^\theta g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta}); h_t). \tag{2}$$

The above preference relation together with (1) also implies

$$g(\sigma(s_i^{\theta'}, s_{-i}^{\theta'}); h_t) \succeq_i^\theta g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta'}); h_t).$$

Since $g(\sigma(s_i^{\theta'}, s_{-i}^{\theta'}); h_t) = g(m_\theta^*; h_t^*) = a$ and we have assumed that θ and θ' are two states satisfying $(*)$, we get that

$$g(\sigma(s_i^{\theta'}, s_{-i}^{\theta'}); h_t) \succeq_i^{\theta'} g(\sigma_i'(s_i^{\theta'}), \sigma_{-i}(s_{-i}^{\theta'}); h_t). \tag{3}$$

Now since by $\mathbf{\Phi2}$, $\phi_i[\cdot \mid s_i^{\theta'}, h_t]$ assigns a strictly positive weight only to $(\theta, s_{-i}^{\theta})$ and $(\theta', s_{-i}^{\theta'})$, Assumption 2 together with (2) and (3) yields:

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i^{\theta'}, h_t]} g((\sigma_i', \sigma_{-i}); h_t).$$

Finally consider the case (3) $h_t \in \mathcal{H}_{-i}^* \backslash \mathcal{H}^*$. Since $h_t \in \mathcal{H}_{-i}^*$ and $h_t \notin \mathcal{H}^*$ (only i has deviated up to t), there must exist $t' < t$ such that $\sigma_i(h_{t'}, s_i^{\theta'}) \neq m_{i,\theta}^*(h_{t'})$ where $h_{t'}$ is a truncation of history h_t. Then, for any history $h_{t''}$ following $h_{t'}$ (and so, in particular, following h_t), we have $h_{t''} \notin \mathcal{H}_{-k}^*$ for each $k \neq i$. Moreover, by $\mathbf{\Sigma1}$ and $\mathbf{\Sigma3}$ we have $\sigma_{-i}(h_{t''}, s_{-i}^{\theta}) = \sigma_{-i}(h_{t''}, s_{-i}^{\theta'}) = m_{-i,\theta}^*(h_{t''})$. Otherwise stated, we have $\sigma_{-i}(s_{-i}^{\theta}) \mid_{h_t} = \sigma_{-i}(s_{-i}^{\theta'}) \mid_{h_t} = m_{-i,\theta}^* \mid_{h_t}$. By $\mathbf{\Phi2}$ we know that $\phi_i[\cdot \mid s_i^{\theta'}, h_t] = \nu^{\varepsilon}(\cdot \mid s_i^{\theta'})$ assigns a strictly positive weight only to $(\theta, s_{-i}^{\theta})$ and $(\theta', s_{-i}^{\theta'})$. In addition, we have that for any $h \in \mathcal{H}^*$ or $h \notin \mathcal{H}_{-i}^* : \sigma_i(h, s_i^{\theta'}) = m_{i,\theta}^*(h, s_i^{\theta'})$. Since $h_t \in \mathcal{H}_{-i}^* \backslash \mathcal{H}^*$, we conclude with $\mathbf{\Sigma2}$

$$g((\sigma_i, \sigma_{-i}); h_t) \succeq_i^{\nu^{\varepsilon}(\cdot|s_i^{\theta'})} g((\sigma_i', \sigma_{-i}); h_t)$$

for any σ_i' that differs from σ_i only at h_t. By Assumption 4, we can apply the one-shot deviation principle, and so the above is equivalent to

$$g((\sigma_i, \sigma_{-i}); h_t) \succeq_i^{\nu^{\varepsilon}(\cdot|s_i^{\theta'})} g((\sigma_i', \sigma_{-i}); h_t)$$

for any σ_i'. This completes the proof. ■

A.2 Consistency

In this section, we show that for some $\phi \in \Phi$, (ϕ, σ) satisfies consistency.

To show this part, we first fix σ as defined above and consider the following sequence $\{(\phi^k, \sigma^k)\}_{k=0}^{\infty}$ of assessments. Let $\eta_k > 0$ for each k and $\eta_k \to 0$ as $k \to \infty$. For each player i, $h_t \in \mathcal{H}$, and signal s_i, let $\xi_i(h_t, s_i, \cdot)$ be any strictly positive prior over $M_i(h_t) \backslash \{\sigma_i(s_i, h_t)\}$ and define σ_i^k as

$$\sigma_i^k(m_i^t \mid h_t, s_i^{\theta'}) = \begin{cases} 1 - \eta_k^{T \times n} \text{ if } m_i^t = \sigma_i(h_t, s_i^{\theta'}); \\ \eta_k^{T \times n} \times \xi_i(h_t, s_i^{\theta'}, m_i^t) \text{ otherwise} \end{cases}$$

where T is the (finite) length of the longest final history; and for any signal $s_i \neq s_i^{\theta'}$:

$$\sigma_i^k(m_i^t \mid h_t, s_i) = \begin{cases} 1 - \eta_k \text{ if } m_i^t = \sigma_i(h_t, s_i); \\ \eta_k \times \xi_i(h_t, s_i, m_i^t) \text{ otherwise} \end{cases}.$$

Let ϕ^k be the unique Bayes consistent belief associated with each σ^k. It is easy to check that σ^k converges to σ and we also have that ϕ^k converges[24]. Let $\phi \equiv \lim_{k\to\infty} \phi^k$. In the sequel, we show that ϕ satisfies **Φ1, Φ2** and **Φ3**. This will show that (ϕ, σ) satisfies consistency, and $\phi \in \Phi$ as claimed.

To do so, we explicitly compute each ϕ^k and study its limit as k tends to infinity. In general for each $(\tilde{\theta}, \tilde{s}_{-i}) \in \Theta \times S_{-i}$, each $h_t = (m^1, ..., m^{t-1}) \in \mathcal{H}$, and each $\tilde{s}_i \in S_i$, we have

$$\phi_i^k[(\tilde{\theta}, \tilde{s}_{-i}) \mid \tilde{s}_i, h_t] = \frac{\nu^\varepsilon(\tilde{\theta}, \tilde{s}_{-i}, \tilde{s}_i) \times \prod_{t'=1}^{t-1} \left[\sigma^k(m^{t'} \mid h_{t'}, \tilde{s})\right]}{\sum\limits_{(\theta', s'_{-i})} \nu^\varepsilon(\theta', s'_{-i}, \tilde{s}_i) \times \prod_{t'=1}^{t-1} \left[\sigma^k(m^{t'} \mid h_{t'}, s'_{-i}, \tilde{s}_i)\right]}.$$

In the above formula for each $t' \leq t$, $h_{t'}$ stands for the truncation of h_t to the first t' elements i.e., $h_{t'} = (m^1, ..., m^{t'-1})$.

Claim 3 *ϕ satisfies* **Φ1.**

Proof of Claim 3: Consider player i, $h_t \notin \mathcal{H}^*_{-i}$. First, we will establish the following lemma.

Lemma 2 *Fix player i and assume that for all $j \neq i$, let $s_j \in \{s_j^\theta, s_j^{\theta'}\}$. Let $h_t = (\emptyset, m^1, ..., m^{t-1}) \notin \mathcal{H}^*_{-i}$.*

(1) There exists $\hat{j} \neq i$ and $\hat{t} \leq t-1$ such that $\sigma_{\hat{j}}(h_{\hat{t}}, s_{\hat{j}}) \neq m_{\hat{j}}^{\hat{t}}$;

(2) If $h_t \in \mathcal{H}^*_{-l}$ for some $l \neq i$, then there exists $\hat{t} \leq t-1$ such that $\sigma_l(h_{\hat{t}}, s_l) \neq m_l^{\hat{t}}$.

Proof of Lemma 2: (1) Assume, on the contrary, that $\sigma_{-i}(h_{t'}, s_{-i}) = m_{-i}^{t'}$ for all $t' \leq t-1$. We then show by induction that for all $t' \leq t$, $h_{t'} \in \mathcal{H}^*_{-i}$, which yields a contradiction. Let $t' = 1$; in this case, $h_1 = \emptyset \in \mathcal{H}^* \subseteq \mathcal{H}^*_{-i}$. Now, toward an induction, assume that $h_{t'-1} \in \mathcal{H}^*_{-i}$ and let us show that $h_{t'} \in \mathcal{H}^*_{-i}$. It is easy to show that $h_{t'-1} \in \mathcal{H}^*_{-i}$ implies that either $h_{t'-1} \in \mathcal{H}^*$ (i.e., no player has deviated) or $h_{t'-1} \notin \mathcal{H}^*_{-j}$ for all $j \neq i$ (i.e., only i has deviated). However, in either case, $\sigma_{-i}(h_{t'-1}, s_{-i}) = m^*_{-i,\theta}(h_{t'-1})$ is obtained by **Σ1** and **Σ3**. Since we have assumed that $\sigma_{-i}(h_{t'-1}, s_{-i}) = m_{-i}^{t'-1}$, we get $m_{-i}^{t'-1} = m^*_{-i,\theta}(h_{t'-1})$, which proves that $h_{t'} = (h_{t'-1}, (\tilde{m}_i(h_{t'-1}), m^*_{-i,\theta}(h_{t'-1}))$ for some strategy $\tilde{m}_i$ and so $h_{t'} \in \mathcal{H}^*_{-i}$. This is a contradiction as desired. (2) Since $h_t \in \mathcal{H}^*_{-l}$, we have that, for all $j \neq l$ and all $t' \leq t-1$, $m_j^{t'} = m^*_{j,\theta}(h_{t'})$. Since $h_t \notin \mathcal{H}^*_{-i}$, we must have

[24] As will become clear from the proof, the sequence $\{\phi^k\}_k$ does converge.

that $h_t \in \mathcal{H}^*_{-l} \backslash \mathcal{H}^*$. Let $\tilde{t} \leq t-1$ be the first date at which $m_l^{\tilde{t}} \neq m^*_{l,\theta}(h_{\tilde{t}})$. By construction, we have that for all $t' \leq \tilde{t}$, $h_{t'} \in \mathcal{H}^*$ while for all $t' > \tilde{t}$, $h_{t'} \notin \mathcal{H}^*_{-j}$ for all $j \neq l$. This implies that for all $j \neq l$ and $t' \leq t-1$, we have $\sigma_j(h_{t'}, s_j) = m^*_{j,\theta}(h_{t'})$ by **Σ1** and **Σ3**. This further implies that for all $j \neq l$ and $t' \leq t-1$, $\sigma_j(h_{t'}, s_j) = m_j^{t'}$. As we already proved in (1), we must have the existence of $\hat{t} \leq t-1$ such that $\sigma_l(h_{\hat{t}}, s_l) \neq m_l^{\hat{t}}$, as claimed. ■

The rest of the proof is reduced to checking the following two cases:

Case 1: $s_i = s_i^{\theta'}$. Recall that $\nu^\varepsilon(\cdot, s_i^{\theta'})$ assigns a weight strictly positive only to $(\theta', s_{-i}^{\theta'})$ and $(\theta, s_{-i}^{\theta})$. Hence,

$$\phi_i^k[(\theta, s_{-i}^\theta) \mid s_i^{\theta'}, h_t]$$

$$= \frac{\nu^\varepsilon(\theta, s_{-i}^\theta, s_i^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta)\right]}{\nu^\varepsilon(\theta, s_{-i}^\theta, s_i^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta)\right] + \nu^\varepsilon(\theta', s_{-i}^{\theta'}, s_i^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})\right]}$$

$$= \frac{\nu^\varepsilon(\theta, s_{-i}^\theta, s_i^{\theta'})}{\nu^\varepsilon(\theta, s_{-i}^\theta, s_i^{\theta'}) + \nu^\varepsilon(\theta', s_{-i}^{\theta'}, s_i^{\theta'}) \times \dfrac{\prod_{j\neq i}[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} | h_{t'}, s_j^{\theta'})]}{\prod_{j\neq i}[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} | h_{t'}, s_j^{\theta})]}}$$

We now show that the ratio $\prod_{j\neq i}[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})] \Big/ \prod_{j\neq i}[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})]$ tends to 0 as k tends to infinity. This will show that $\phi_i^k[(\theta, s_{-i}^\theta) \mid s_i^{\theta'}, h_t] \to 1$ and $\phi_i^k[(\theta', s_{-i}^{\theta'}) \mid s_i^{\theta'}, h_t] \to 0$.

By construction of σ^k, Lemma 2 (1) implies that for some $\hat{j} \neq i$ and $\hat{t} \leq t-1:$

$$\sigma_{\hat{j}}^k(m_{\hat{j}}^{\hat{t}} \mid h_{\hat{t}}, s_{\hat{j}}^{\theta'}) = \eta_k^{T\times n} \xi_{\hat{j}}(h_{\hat{t}}, s_{\hat{j}}^{\theta'}, m_{\hat{j}}^{\hat{t}}). \tag{4}$$

Now, we have:

$$\frac{\prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})\right]}{\prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})\right]} \leq \frac{\eta_k^{T\times n} \times \xi_{\hat{j}}(h_{\hat{t}}, s_{\hat{j}}^{\theta'}, m_{\hat{j}}^{\hat{t}}) \times 1}{\prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \eta_k \xi_j(h_{t'}, s_j^\theta, m_j^{t'})\right]}$$

$$= \frac{\eta_k^{T\times n}}{\eta_k^{(t-1)(n-1)}} \times \frac{\xi_{\hat{j}}(h_{\hat{t}}, s_{\hat{j}}^{\theta'}, m_{\hat{j}}^{\hat{t}})}{\prod_{j\neq i}\left[\prod_{t'=1}^{t-1} \xi_j(h_{t'}, s_j^\theta, m_j^{t'})\right]} \to 0 \text{ (as } k \to \infty).$$

Where the first inequality is assured by (4) and (assuming wlog that η_k is small) we use the very construction that, for all j and $t' \leq t-1$, $\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta) \geq \eta_k \times \xi_j(h_{t'}, s_j^\theta, m_j^{t'})$.

Case 2: $s_i = s_i^\theta$. Recall that $\nu^\varepsilon(\cdot, s_i^\theta)$ assigns a weight strictly positive only to (θ, s_{-i}^θ) and (θ, τ_l) for each $l \neq i$. Hence,

$$\begin{aligned}
&\phi_i^k[(\theta,\tau_l) \mid s_i^\theta, h_t] \\
&= \frac{\nu^\varepsilon(\theta,\tau_l) \times \prod_{j\neq l,i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta)\right] \times \left[\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})\right]}{\left\{\begin{array}{c}\sum_{z\neq i}\nu^\varepsilon(\theta,\tau_z) \times \prod_{j\neq z,i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta)\right] \times \left[\prod_{t'=1}^{t-1}\sigma_z^k(m_z^{t'} \mid h_{t'}, s_z^{\theta'})\right] \\ +\nu^\varepsilon(\theta, s_{-i}^\theta, s_i^\theta) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^\theta)\right]\end{array}\right\}} \\
&= \frac{\nu^\varepsilon(\theta,\tau_l)}{\sum_{z\neq i}\nu^\varepsilon(\theta,\tau_z) \times c_z(k) + \nu^\varepsilon(\theta, s_{-i}^\theta, s_i^\theta) \times \frac{\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'}|h_{t'},s_l^\theta)}{\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'}|h_{t'},s_l^{\theta'})}}
\end{aligned}$$

for some positive functions $c_z(k)$. We now show that if $h_t \in \mathcal{H}_{-l}^*$, then the ratio $\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'}, h_{t'}, s_l^\theta) \big/ \prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})$ tends to ∞ as k tends to infinity. This will show that $\phi_i^k[(\theta,\tau_l) \mid s_i^\theta, h_t] \to 0$ for all l such that $h_t \in \mathcal{H}_{-l}^*$; and hence that ϕ satisfies **Φ1.** Assume that $h_t \in \mathcal{H}_{-l}^*$ for some l, by construction of σ^k, Lemma 2 (2) implies that there exists $\hat{t} \leq t-1$ such that $\sigma_l(h_{\hat{t}}, s_l) \neq m_l^{\hat{t}}$ and so:

$$\sigma_l^k(m_l^{\hat{t}} \mid h_{\hat{t}}, s_l^{\theta'}) = \eta_k^{T\times n}\xi_l(h_{\hat{t}}, s_l^{\theta'}, m_l^{\hat{t}}). \tag{5}$$

Now, we have

$$\frac{\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^\theta)}{\prod_{t'=1}^{t-1}\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})} \quad \geq \quad \frac{\eta_k^{t-1}\prod_{t'=1}^{t-1}\xi_l(h_{t'}, s_l^\theta, m_l^{t'})}{\eta_k^{T\times n}\xi_l(h_{\hat{t}}, s_l^{\theta'}, m_l^{\hat{t}}) \times 1} \to \infty \text{ (as } k \to \infty).$$

Where the first inequality is assured by (5) and (assuming wlog that η_k is small) we use the fact that by construction: for all $t' \leq t-1 : \sigma_l^k(m_j^{t'} \mid h_{t'}, s_l^\theta) \geq \eta_k \times \xi_l(h_{t'}, s_l^\theta, m_l^{t'})$. ■

Claim 4 *ϕ satisfies* **Φ2.**

Proof of Claim 4: Consider player i, $h_t \in \mathcal{H}^*_{-i}$. The following lemma will be useful.

Lemma 3 *Fix player i and assume that for all $j \neq i$, $s_j \in \{s_j^\theta, s_j^{\theta'}\}$. Let $h_t = (\emptyset, m^1, ..., m^{t-1}) \in \mathcal{H}^*_{-i}$. For all $j \neq i$ and $t' \leq t-1 : \sigma_j(h_{t'}, s_j) = m_j^{t'}$.*

Proof of Lemma 3: Pick any $t' \leq t-1$ and note that $h_{t'} \in \mathcal{H}^*_{-i}$. Hence, it must be that either $h_{t'} \in \mathcal{H}^*$ or $h_{t'} \notin \mathcal{H}^*_{-j}$ for all $j \neq i$. In each of these cases, by **Σ1** and **Σ3**, we have for all $j \neq i : \sigma_j(h_{t'}, s_j) = m^*_{j,\theta}(h_{t'})$. Since $h_{t'} \in \mathcal{H}^*_{-i}$, we have that, for all $j \neq i$, $m_j^{t'} = m^*_{j,\theta}(h_{t'})$, which completes the proof. ■

Here again, the rest of the proof is reduced to checking the following two cases.

Case 1: $s_i = s_i^{\theta'}$. Recall that $\nu^\varepsilon(\cdot, s_i^{\theta'})$ assigns a weight strictly positive only to $(\theta', s_{-i}^{\theta'})$ and $(\theta, s_{-i}^{\theta})$. Hence,

$$\begin{aligned}
&\phi_i^k[(\theta', s_{-i}^{\theta'}) \mid s_i^{\theta'}, h_t] \\
&= \frac{\nu^\varepsilon(\theta', s_{-i}^{\theta'}, s_i^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})\right]}{\nu^\varepsilon(\theta', s^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})\right] + \nu^\varepsilon(\theta, s_{-i}^{\theta}, s_i^{\theta'}) \times \prod_{j\neq i}\left[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})\right]} \\
&= \frac{\nu^\varepsilon(\theta', s_{-i}^{\theta'}, s_i^{\theta'})}{\nu^\varepsilon(\theta', s^{\theta'}) + \nu^\varepsilon(\theta, s_{-i}^{\theta}, s_i^{\theta'}) \times \frac{\prod_{j\neq i}[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'}|h_{t'}, s_j^{\theta})]}{\prod_{j\neq i}[\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'}|h_{t'}, s_j^{\theta'})]}}
\end{aligned}$$

We now show that the ratio $\prod_{j\neq i}\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta}) / \prod_{j\neq i}\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'})$ tends to 1 as k tends to infinity. This will show that $\phi_i^k[(\theta', s_{-i}^{\theta'}) \mid s_i^{\theta'}, h_t] \to \nu^\varepsilon((\theta', s_{-i}^{\theta'}) \mid s_i^{\theta'})$ and $\phi_i^k[(\theta, s_{-i}^{\theta}) \mid s_i^{\theta'}, h_t] \to \nu^\varepsilon((\theta, s_{-i}^{\theta}) \mid s_i^{\theta'})$.

By construction of σ^k, Lemma 3 implies that for all $j \neq i$ and $t' \leq t-1$:

$$\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta}) = 1 - \eta_k \text{ and } \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'}) = 1 - \eta_k^{T\times n}$$

Thus,

$$\prod_{j\neq i}\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta}) \Big/ \prod_{j\neq i}\prod_{t'=1}^{t-1}\sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta'}) \to 1 \text{ as } k \to \infty$$

Case 2: $s_i = s_i^{\theta}$. Recall that $\nu^\varepsilon(\cdot, s_i^{\theta})$ assigns a weight strictly positive only to

$(\theta, s_{-i}^{\theta})$ and (θ, τ_l) for $l \neq i$. Hence,

$$\phi_i^k[(\theta, s_{-i}^{\theta}) \mid s_i^{\theta}, h_t]$$

$$= \frac{\nu^{\varepsilon}(\theta, s_{-i}^{\theta}, s_i^{\theta}) \times \prod_{j \neq i} \left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})\right]}{\left\{\begin{array}{c} \nu^{\varepsilon}(\theta, s_{-i}^{\theta}, s_i^{\theta}) \times \prod_{j \neq i} \left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})\right] \\ +\sum_{l \neq i} \nu^{\varepsilon}(\theta, \tau_l) \times \prod_{j \neq i,l} \left[\prod_{t'=1}^{t-1} \sigma_j^k(m_j^{t'} \mid h_{t'}, s_j^{\theta})\right] \times \left[\prod_{t'=1}^{t-1} \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})\right] \end{array}\right\}}$$

$$= \frac{\nu^{\varepsilon}(\theta, s_{-i}^{\theta}, s_i^{\theta})}{\nu^{\varepsilon}(\theta, s_{-i}^{\theta}, s_i^{\theta}) + \sum_{l \neq i} \nu^{\varepsilon}(\theta, \tau_l) \times \frac{\prod_{t'=1}^{t-1} \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})}{\prod_{t'=1}^{t-1} \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta})}}$$

We now show that for each $l \neq i$, the ratio $\prod_{t'=1}^{t-1}[\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'})] \Big/ \prod_{t'=1}^{t-1}[\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta})]$ tends to 1 as k tends to infinity. This will show that $\phi_i^k[(\theta, s_{-i}^{\theta}) \mid s_i^{\theta}, h_t] \to \nu^{\varepsilon}((\theta, s_{-i}^{\theta}) \mid s_i^{\theta})$ and similar reasoning shows that for each $l \neq i$: $\phi_i^k[(\theta, \tau_l) \mid s_i^{\theta}, h_t] \to \nu^{\varepsilon}((\theta, \tau_l) \mid s_i^{\theta})$; and hence, ϕ satisfies **Φ2.**

By construction of σ^k, Lemma 3 implies that for all $l \neq i$ and $t' \leq t-1$:

$$\sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta}) = 1 - \eta_k \text{ and } \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'}) = 1 - \eta_k^{T \times n}$$

Thus,

$$\prod_{t'=1}^{t-1} \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta'}) \Big/ \prod_{t'=1}^{t-1} \sigma_l^k(m_l^{t'} \mid h_{t'}, s_l^{\theta}) \to 1 \text{ as } k \to \infty$$

■

Finally, observing that for $s_i^{\tilde{\theta}} \notin \{s_i^{\theta}, s_i^{\theta'}\}$, $\nu^{\varepsilon}(\cdot, s_i^{\tilde{\theta}})$ assigns a weight one to $(\tilde{\theta}, s_{-i}^{\tilde{\theta}})$, we establish the claim below:

Claim 5 *ϕ satisfies* **Φ3.**

B One-Shot Deviation Principle

As already mentioned, Hendon, Jacobsen and Sloth (1996) proved the one-shot deviation principle for sequential equilibrium when players are expected utility decision makers. In this section, we show that in our setting where agents need not be expected utility decision makers, under two weak additional assumptions (obviously satisfied by expected utility models), the one-shot deviation principle holds.

To state the assumptions, assume that one dimension – say Y – is added to the state space. We will have to define order of preferences over this extended domain. We say that α_Y is a Y-act if α_Y is a mapping from $\Theta \times S \times Y$ to A. We assume that for each belief $\beta \in \Delta(\Theta \times S \times Y)$, each player i has a *transitive* preference relation $\succeq_i^\beta$ over Y-acts.

We make two weak assumptions that hold in the expected utility case. The first assumption is in the spirit of Assumption 2 in the paper and imposes a weak restriction on preference orders when the domain of acts is extended from $\Theta \times S$ to $\Theta \times S \times Y$.

Assumption 5 *Fix any countable set Y and take two Y-acts α_Y and $\hat{\alpha}_Y$, and a belief $\beta \in \Delta(\Theta \times S \times Y)$. We have*

$$\left(\alpha(\cdot,\cdot,y) \succeq_i^{\beta(\cdot,\cdot|y)} \hat{\alpha}(\cdot,\cdot,y) \text{ for all } y \text{ s.t. } \beta(y) > 0\right) \Rightarrow \alpha \succeq_i^\beta \hat{\alpha}.$$

The second assumption is also very weak. First, we say that a Y-act α_Y coincides with an act α under $\beta \in \Delta(\Theta \times S \times Y)$ if for any $\theta, s, y : \beta(\theta, s, y) > 0 \Rightarrow \alpha_Y(\theta, s, y) = \alpha(\theta, s)$. The following assumption only assumes that if under β, α_Y and $\hat{\alpha}_Y$ respectively coincide with acts α and $\hat{\alpha}$, then the uncertainty dimension Y can be ignored. More precisely:

Assumption 6 *Fix any countable set Y and take two Y-acts α_Y and $\hat{\alpha}_Y$, and a belief $\beta \in \Delta(\Theta \times S \times Y)$. Assume that α_Y and $\hat{\alpha}_Y$ respectively coincide with acts α and $\hat{\alpha}$ under β. We have*

$$\alpha_Y \succeq_i^\beta \hat{\alpha}_Y \Rightarrow \alpha \succeq_i^{marg_{\Theta \times S}\ \beta} \hat{\alpha}.$$

For simplicity, we assume without loss of generality that (1) for all player i and $h \notin H_T : |M_i(h)| \geq 1$ and (2) any terminal histories $h, h' \in H_T$ have the same length T. [25]

[25] This is without loss of generality because assuming that player i does not play after history h (i.e. assuming that $M_i(h) = \emptyset$), or assuming that player i has a unique available message after history h, (i.e. $|M_i(h)| = 1$) is essentially the same: this does not affect the sequential equilibrium outcomes. Hence, assuming that all terminal histories have the same length T is also without loss of generality.

Proposition 1 (One-Shot Deviation Principle) *Suppose Assumptions 5 and 6 hold. Let* (ϕ, σ) *be an assessment that satisfies consistency. For each* i, s_i, t *and* h_t :

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for every σ_i' *if and only if for each* i, s_i, t *and* h_t :

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for every σ_i' *that differs from* σ_i *only at* h_t.

Proof of Proposition 1.

We will first need the following claim.

Claim 6 *Let* $h_{t+1} = (h_t, m^t)$ *where* $m^t \in M(h_t)$. *We have for each* i *and* s_i

$$\phi_i(\theta, s_{-i} \mid s_i, h_t)\sigma(m^t \mid h_t, s_i, s_{-i}) = \left(\sum_{\tilde{s}_{-i}} \phi_i[\tilde{s}_{-i} \mid s_i, h_t]\sigma(m^t \mid h_t, s_i, \tilde{s}_{-i})\right) \phi_i(\theta, s_{-i} \mid s_i, h_{t+1}).$$

Proof. In case $\sum_{\tilde{s}_{-i}} \phi_i[\tilde{s}_{-i} \mid s_i, h_t]\sigma(m^t \mid h_t, s_i, \tilde{s}_{-i}) = 0$ then the right-hand side is zero but obviously so is the left-hand side. Now assume $\sum_{\tilde{s}_{-i}} \phi_i[\tilde{s}_{-i} \mid s_i, h_t]\sigma(m^t \mid h_t, s_i, \tilde{s}_{-i}) > 0$. By consistency, we can pick the sequence of totally mixed strategy profile $\sigma^k \to \sigma$ such that (recall that for each $t' \leq t$, $h_{t'} = (m^1, m^2, ..., m^{t'-1})$ stands for the truncation of h_t to the first t' elements)

$$\begin{aligned}
&\frac{\phi_i(\theta, s_{-i} \mid h_t, s_i)\sigma(m^t \mid h_t, s_i, s_{-i})}{\sum_{\tilde{s}_{-i}} \phi_i[\tilde{s}_{-i} \mid s_i, h_t]\sigma(m^t \mid h_t, s_i, \tilde{s}_{-i})} \\
&= \lim_{k\to\infty} \frac{\nu(\theta, s_{-i}, s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, s_{-i})}{\sum_{\theta, s_{-i}'} \nu(\theta, s_{-i}', s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, s_{-i}')} \times \sigma_{-i}(m_{-i}^t \mid h_t, s_{-i}) \\
&\Bigg/ \lim_{k\to\infty} \sum_{\tilde{s}_{-i}} \frac{\nu(\tilde{s}_{-i}, s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, \tilde{s}_{-i})}{\sum_{s_{-i}'} \nu(s_{-i}', s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, s_{-i}')} \times \sigma_{-i}(m_{-i}^t \mid h_t, \tilde{s}_{-i})
\end{aligned}$$

$$= \lim_{k\to\infty} \frac{\nu(\theta, s_{-i}, s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, s_{-i}) \times \sigma_{-i}(m_{-i}^t \mid h_t, s_{-i})}{\sum_{\tilde{s}_{-i}} \nu(\tilde{s}_{-i}, s_i) \times \prod_{t'=1}^{t-1} \sigma_{-i}^k(m_{-i}^{t'} \mid h_{t'}, \tilde{s}_{-i}) \times \sigma_{-i}(m_{-i}^t \mid h_t, \tilde{s}_{-i})}$$

$$= \phi_i(\theta, s_{-i} \mid h_{t+1}, s_i)$$

as claimed. ■

We are now in a position to prove our Proposition 1.

Assume (ϕ, σ) satisfies local sequential rationality, i.e. for each i, s_i, t and h_t :

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for every σ_i' that differs from σ_i only at h_t.

Fix any i, s_i. Recall that T is the length of any terminal history. We want to show by induction on k that for any $k \geq 1$ and any $t \geq T - k$, any h_t satisfies:

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t)$$

for any strategy σ_i'. First, note that this is true for $k = 1$ because (ϕ, σ) satisfies local sequential rationality. Now toward an induction, assume (IH) that for any $t' \geq T - k$, any $h_{t'}$ satisfies

$$g(\sigma; h_{t'}) \succeq_i^{\phi_i[\cdot|s_i,h_{t'}]} g((\sigma_i', \sigma_{-i}); h_{t'})$$

for any strategy σ_i'. Pick any h_t such that $t \geq T - (k + 1)$ and fix any strategy σ_i'. By local sequential rationality, we know that

$$g(\sigma; h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\hat{\sigma}_i, \sigma_{-i}); h_t)$$

for $\hat{\sigma}_i(h) = \begin{cases} \sigma_i'(h) \text{ if } h = h_t \\ \sigma_i(h) \text{ otherwise} \end{cases}$; hence, by transitivity of $\succeq_i^{\phi_i[\cdot|s_i,h_t]}$, it is enough to show that

$$g((\hat{\sigma}_i, \sigma_{-i}); h_t) \succeq_i^{\phi_i[\cdot|s_i,h_t]} g((\sigma_i', \sigma_{-i}); h_t). \tag{6}$$

For any profile of strategies σ, let the $M(h_t)$-act $\tilde{g}(\sigma; h_t) : \Theta \times S \times M(h_t) \to A$ be such that $(\theta, s, m^t) \mapsto g(\sigma_i(s_i), \sigma_{-i}(s_{-i}); h_t, m^t)$; and let $\tilde{\phi}_i[\cdot \mid s_i, h_t]$ be the distribution over $\Theta \times S \times M(h_t)$ such that

$$\tilde{\phi}_i[(\theta, s, m^t) \mid s_i, h_t] = \phi_i[(\theta, s_{-i}) \mid s_i, h_t]\sigma_i'(m_i^t \mid h_t, s_i)\sigma_{-i}(m_{-i}^t \mid h_t, s_{-i}).$$

Note that $\tilde{\phi}_i[(\theta, s, m^t) \mid s_i, h_t] > 0$ implies that[26] $\sigma_i'(s_i, h_t) = m_i^t$ and $\sigma_{-i}(s_{-i}, h_t) = m_{-i}^t$. Hence, whenever $\tilde{\phi}_i[(\theta, s, m^t) \mid s_i, h_t] > 0$:

$$\begin{aligned}\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t)(\theta, s, m^t) &= g(\hat{\sigma}_i(s_i), \sigma_{-i}(s_{-i}); h_t, \sigma_i'(h_t, s_i), \sigma_{-i}(h_t, s_{-i})) \\ &= g(\hat{\sigma}_i(s_i), \sigma_{-i}(s_{-i}); h_t)\end{aligned}$$

where the first equality is by definition of $\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t)$ while the second is by definition of $\hat{\sigma}_i$. Thus, the $M(h_t)$-act $\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t)$ coincides with the act $g((\hat{\sigma}_i, \sigma_{-i}); h_t)$ under $\tilde{\phi}_i[\cdot \mid s_i, h_t]$. Similarly, one can easily check that the $M(h_t)$-act $\tilde{g}((\sigma_i', \sigma_{-i}); h_t)$ coincides with the act $g((\sigma_i', \sigma_{-i}); h_t)$ under $\tilde{\phi}_i[\cdot \mid s_i, h_t]$. Now, Assumption 6 together with the fact that $\phi[\cdot \mid s_i, h_t] = \text{marg}_{\Theta \times S_{-i}} \tilde{\phi}[\cdot \mid s_i, h_t]$ implies that (6) holds if

$$\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t) \succeq_i^{\tilde{\phi}_i[\cdot|s_i, h_t]} \tilde{g}((\sigma_i', \sigma_{-i}); h_t).$$

By Claim 6, we know that for any m^t :

$$\tilde{\phi}_i[(\theta, s, m^t) \mid s_i, h_t] = \left(\sum_{\tilde{s}_{-i}} \phi_i[\tilde{s}_{-i} \mid s_i, h_t] \sigma_i'(m_i^t \mid h_t, s_i) \sigma_{-i}(m_{-i}^t \mid h_t, \tilde{s}_{-i}) \right) \phi_i(\theta, s_{-i} \mid s_i, h_t, m^t). \tag{7}$$

In addition, because h_t has been chosen so that $t + 1 \geq T - k$, the inductive hypothesis (IH) applies to $h_{t+1} = (h_t, m^t)$ i.e.

$$g((\sigma_i, \sigma_{-i}); h_t, m^t) \succeq_i^{\phi_i[\cdot|s_i, h_t, m^t]} g((\sigma_i', \sigma_{-i}); h_t, m^t). \tag{8}$$

Now, it is clear that $g((\sigma_i', \sigma_{-i}); h_t, m^t) = \tilde{g}((\sigma_i', \sigma_{-i}); h_t)(\cdot, \cdot, m^t)$; in addition, by definition of $\hat{\sigma}_i$, $\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t)(\cdot, \cdot, m^t) = g((\sigma_i, \sigma_{-i}); h_t, m^t)$. Thus, by (8), we get that for any $m^t \in M(h_t)$:

$$\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t)(\cdot, \cdot, m^t) \succeq_i^{\phi_i[\cdot|s_i, h_t, m^t]} \tilde{g}((\sigma_i', \sigma_{-i}); h_t)(\cdot, \cdot, m^t) \tag{9}$$

Therefore, by (7), it is easily checked that $\phi[\cdot \mid s_i, h_t]$ is equal to $\tilde{\phi}[\cdot \mid s_i, h_t]$ conditional on m^t. Hence, using (9) and Assumption 5, we get that

$$\tilde{g}((\hat{\sigma}_i, \sigma_{-i}); h_t) \succeq_i^{\tilde{\phi}_i[\cdot|s_i, h_t]} \tilde{g}((\sigma_i', \sigma_{-i}); h_t)$$

which completes the proof. ■

[26] Recall that we consider only pure strategies.

References

[1] D. Abreu, and H. Matsushima, (1992) "Virtual Implementation in Iteratively Undominated Strategies: Complete Information", *Econometrica*, 60, 993-1008

[2] D. Abreu and A. Sen (1990) "Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition", *Journal of Economic Theory*, 50, 285-299

[3] Aghion P., D. Fudenberg, R. and Holden (2009) "Subgame Perfect Implementation With Almost Perfect Information and the Hold-Up Problem," working paper, available at http://www.economics.harvard.edu/faculty/aghion/files/Subgame_Perfect.pdf

[4] Che, Y.-K. and D. Hausch. (1999) "Cooperative Investments and the Value of Contracting," *American Economic Review*, 89, 125-147

[5] Chung, K. and J. Ely (2003) "Implementation with Near-Complete Information," *Econometrica*, 71, 857-87

[6] J. Duggan (1997) "Virtual Bayesian Implementation," *Econometrica*, 65, 1175-1199

[7] D. Fudenberg, D.M. Kreps, and D.K. Levine (1988) "On the Robustness of Equilibrium Refinements," *Journal of Economic Theory*, 44, 354-380

[8] D. Fudenberg and J. Tirole (1991) "*Game Theory*," MIT Press

[9] J. Glazer and R.W. Rosenthal (1992) "A Note on Abreu-Matsushima Mechanisms," *Econometrica*, 60, 1435-1438

[10] S.J. Grossman and O.D. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691-719

[11] O. Hart and J. Moore (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119-1158

[12] Hendon, E., H.J. Jacobsen and B. Sloth, (1996) "The One-Shot Deviation Principle for Sequential Rationality," *Games and Economic Behavior*, 12, 274-282

[13] Kreps, D.M. and R. Wilson (1982) "Sequential Equilibria," *Econometrica*, 50, 863-894

[14] Kunimoto, T (2008) "How Robust is Undominated Nash Implementation?," *mimeo* available at http://people.mcgill.ca/files/takashi.kunimoto/robustness-Dec-9-2008.pdf

[15] Kunimoto, T (2009) "Robust Virtual Implementation with Almost Complete Inforamtion," *mimeo* available at http://people.mcgill.ca/files/takashi.kunimoto/RVI-almost-complete-May-21-2009.pdf

[16] Maskin, E. (1999) "Nash Equilibirum and Welfare Optimality," *Review of Economic Studies*, 66, 23-38

[17] Maskin, E. and J. Tirole (1999a) "Unforseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, 66, 83-114

[18] Maskin, E. and J. Tirole (1999b) "Two Remarks on the Property-Rights Literature," *Review of Economic Studies*, 66, 139-149

[19] E. Miyagawa (2002) "Subgame-Perfect Implementation of Bargining Solutions," *Games and Economic Behavior*, 41, 292-302

[20] Monderer, D. and D. Samet (1989) "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1, 170-190

[21] Moore, J. and R. Repullo (1988), "Subgame Perfect Implementation," *Econometrica*, 56, 1191-1220

[22] Moore, J. (1992), "Implementation, Contracts and Renegotiation in Environments with Complete Information," in J-J Laffont (ed.), Advances in Economic Theory: Invited Papers to the World Congress of the Econometric Society, vol.I, 182-282

[23] Oury, M. and O. Tercieux (2009), "Continuous Implementation," working paper, available at http://www.pse.ens.fr/tercieux/ContUTP.pdf

[24] Palfrey, T.R., and S. Srivastava (1991), "Nash Implementation Using Undominated Strategies," *Econometrica*, 59, 479-501

[25] Sefton, M. and A. Yavas (1996), "Abreu-Matsushima Mechanisms: Experimental Evidence," *Games and Economic Behavior*, 16, 280-302

[26] H. Vartiainen (2007), "Subgame Perfect Implementation: A Full Characterization," *Journal of Economic Theory*, 133, 111-126