Robust Virtual Implementation under Common Strong Belief in Rationality†

Christoph Müller‡
University of Minnesota
JOB MARKET PAPER
November 17, 2009

Abstract

Robust virtual implementation asks if a social goal can be approximately achieved if merely the agents’ rationality is common knowledge. Bergemann and Morris (2009b) show that static mechanisms can robustly virtually implement essentially no social goal if preferences are sufficiently interdependent. Without any knowledge of how agents revise their beliefs this impossibility result extends to dynamic mechanisms, and focusing on static mechanisms is without loss of generality. In contrast, this paper shows that excluding dynamic mechanisms entails considerable loss of generality if agents commonly believe in rationality “as long as possible”. We illustrate this in private consumption environments with discrete payoff types and generic valuation functions. In such environments, dynamic mechanisms can robustly virtually implement all ex-post incentive compatible social goals regardless of the level of preference interdependence. This result derives from the key insight that under common strong belief in rationality (Battigalli and Siniscalchi, 2002), dynamic mechanisms can almost always distinguish all payoff type profiles by their strategic choices. Notably, dynamic mechanisms can robustly virtually implement the efficient allocation of an object even if static mechanisms cannot.

†I am indebted to my advisor Kim-Sau Chung for his dedication and encouragement, and for his invaluable guidance throughout this project. I thank David Rahman, Itai Sher and Jan Werner for their continuous support, and Dirk Bergemann, Narayana Kocherlakota, Eric Maskin, Konrad Mierendorff, Stephen Morris, Guillermo Ordonez, Alessandro Pavan and Christopher Phelan for helpful discussions and suggestions. I also thank the participants of the Workshop on Information and Dynamic Mechanism Design (HIM, Bonn), the SED Annual Meetings (Istanbul), the ESEM (Barcelona) and the Mathematical Economics Workshop at the University of Minnesota. Financial support through a Doctoral Dissertation Fellowship of the Graduate School of the University of Minnesota is gratefully acknowledged. All errors are my own.

‡Contact: christmu@econ.umn.edu
1 Introduction

Recently, Bergemann and Morris (2009b) (henceforth BM) introduced the concept of strategic distinguishability as a multi-person analogue to the single-person revealed preference theory. In their paper, two payoff types of an agent are called strategically distinguishable if there exists a static mechanism (normal form game form) in which these two payoff types have disjoint sets of rationalizable strategies. BM’s focus on rationalizable strategies is motivated by their concern for robustness, which keeps them from assuming that anything beyond the agents’ rationality is common knowledge.† But BM’s focus on static mechanisms does not have a similar motivation. Their notion of strategic distinguishability generalizes without difficulty to dynamic mechanisms (roughly, extensive form game forms), and one may wonder what one may have lost by excluding the latter.

We show that the answer to this question crucially depends on what is known about how agents revise their beliefs at information sets that they did not expect to occur. Notice that the decision of what belief-revision assumptions to make does not arise in static mechanisms, but becomes unavoidable in dynamic mechanisms. In Müller (2009) we prove that, if we do not impose any belief-revision assumptions, focusing on static mechanism is indeed without loss of generality. In that case, dynamic mechanisms cannot strategically distinguish more payoff types than static mechanisms.

In this paper, we show that if we are willing to impose one belief-revision assumption, namely common strong belief in rationality (Battigalli and Siniscalchi, 2002), then focusing on static mechanisms is with considerable loss of generality. We illustrate this in the context of private consumption environments. BM show how to strategically distinguish agents’ payoff types if there is little preference interdependence. But BM also point out that if preferences are sufficiently interdependent, it is impossible to find a static mechanism that strategically distinguishes at least some payoff types of some agent. In contrast, all payoff types of all agents can be strategically distinguished by dynamic mechanisms in private consumption environments with generic valuation functions, and therefore (essentially) regardless of the degree of preference interdependence (proposition 3).

This result has significant practical implications. BM prove that ex-post incentive compatibility (epIC) and robust measurability are necessary and, under an economic assumption, sufficient for robust virtual implementation in static mechanisms. A social choice function is robustly measurable if it treats any payoff types the same that are strategically indistinguishable by static mechanisms. We show that if we appropriately generalize robust measurability, analogous necessary (proposition 1) and, with minor qualifications, analogous sufficient conditions (proposition 2) hold if we allow for dynamic mechanisms. Hence the fact that dynamic

†Dekel, Fudenberg, and Morris (2007) prove that a strategy is consistent with rationality and common belief in rationality precisely if it is (interim correlated) rationalizable.
mechanisms can strategically distinguish more payoff types immediately implies that they can also robustly virtually implement more social choice functions. In particular, in private-consumption environments, epiC social choice functions can be robustly virtually implemented in dynamic mechanisms (essentially) regardless of the degree of preference interdependence, but robustly virtually implemented in static mechanisms only when there is sufficiently little preference interdependence (or if the social choice function is constant).

1.1 Preview of Results

Subsubsections 1.1.1 and 1.1.2 describe the results we obtain if agents are rational and if there is common strong belief in rationality (RCSBR) in more detail. An agent strongly believes in an event if he initially believes in the event, and continues to believe in the event “as long as possible”. That an agent strongly believes in an event says something about how he revises his belief. For example, if i strongly believes in j’s rationality then i believes that j is rational at all information sets, including those that he did not expect to occur, save for those that can have resulted only from irrational play of j. Subsubsection 1.1.3 contrasts these results with results for the case that only static mechanisms are available and the case that there are no belief-revision assumptions in place. The latter case arises if we only assume that agents are rational and that there is common initial belief in rationality (RCIBR).

In our analysis, we restrict attention to belief-complete type spaces. This allows us to use a result by Battigalli and Siniscalchi (2002), which shows that in such type spaces, RCSBR is characterized by strong rationalizability (Battigalli, 2003). Furthermore, we rule out badly behaved mechanisms by restricting attention to finite dynamic mechanisms, and focus on finite payoff type spaces. Finiteness is an attractive feature of mechanisms, and finiteness of payoff type spaces a standard assumption in the virtual implementation literature†.

1.1.1 Strategic Distinguishability

We call two payoff type profiles strategically indistinguishable if in any dynamic mechanism some terminal node is reached by strongly rationalizable strategy profiles of both payoff type profiles (and strategically distinguishable otherwise). If such a terminal node occurs, it is impossible to tell which (if any) of the two payoff type profiles has played the mechanism. If in this definition we replace “strongly rationalizable” with “weakly rationalizable”, we obtain the notion of strategic indistinguishability we use in Müller (2009). Weak rationalizability (Battigalli, 2003) is characterized by RCIBR and incorporates no belief-revision assumptions. If instead we replace “dynamic mechanism” with “static mechanism” we obtain the original notion of strategic distinguishability introduced by BM (reformulated for payoff type profiles).

†Compare BM; Abreu and Matsushima (1992a,b); Artemov, Kunimoto, and Serrano (2009). BM provide an example of strategic distinguishability with continuous payoff type spaces.
Section 5 examines strategic distinguishability (under RCSBR) in environments with interdependent preferences† in which an object is to be allocated between a finite number of agents, utilities are quasilinear and lotteries over allocations are available (private consumption environments). In private consumption environments, an ex-post valuation of a payoff type is any valuation the payoff type can have for the object if he has a degenerate belief about the other agents’ payoff types. Proposition 3 shows that if for every agent, the sets of ex-post valuations are disjoint for any two distinct payoff types, then all payoff type profiles are strategically distinguishable. The mechanism that strategically distinguishes all payoff type profiles has a simple structure: One after another, agents announce a possible ex-post valuation. Each agent moves only once.

The sufficient condition of proposition 3 holds for almost all valuation function profiles. It merely requires that the set of ex-post valuations be disjoint across an agent’s payoff types — distinct payoff types can have the same valuation for non-degenerate beliefs. Moreover, the sufficient condition is not even necessary (example 5.1): Even if for all payoff types of all agents the sets of ex-post valuations intersect, it is possible that all payoff type profiles are strategically distinguishable. Yet, it is worth noting that it is not always possible to strategically distinguish all payoff type profiles. This follows from proposition 4, which provides a weak necessary condition for the strategic distinguishability of payoff types, and therefore, of payoff type profiles‡: If a valuation is an ex-post valuation of two payoff types for the same degenerate belief about others’ payoff types, then the two payoff types cannot be strategically distinguished.

1.1.2 Robust Virtual Implementation

Our results on strategic distinguishability have a direct application in robust virtual implementation. We say that a dynamic mechanism robustly implements a social choice function if for every payoff type profile, every strongly rationalizable strategy profile induces the outcome prescribed by the social choice function. That is, a social choice function is robustly implementable if it is fully implementable under strong rationalizability. A social choice function is robustly virtually implementable (rv-implementable) roughly if arbitrarily close-by social choice functions are robustly implementable. Analogously to above, replacing “strongly rationalizable” with “weakly rationalizable” yields the implementation concept we study in Müller (2009), and replacing “dynamic mechanism” with “static mechanism” yields the implementation concept studied by BM. Note that mechanisms that are robust in the sense of any of

†Preferences are interdependent if an agent’s preferences not only depend on his own payoff type, but also on the payoff types of others.

‡Two payoff types are strategically indistinguishable if in any mechanism, strongly rationalizable strategies of both of the payoff types together with some strongly rationalizable strategy profile of a fixed payoff type profile of the others’ can lead to the same terminal node.
these definitions do not rely on common knowledge assumptions about agents’ initial beliefs and higher order beliefs about others’ payoff types.

Section 4 examines rv-implementation in general environments in which outcomes are lotteries over a finite set of pure outcomes and agents have expected utility preferences. Slightly weakening the implementation concept from robust to robust virtual implementation leads to a simple relation between implementability and strategic distinguishability (propositions 1 and 2). This relation is a “descendant” of Abreu and Matsushima’s (1992b) characterization of (non-robust) virtual implementation in static mechanisms.

Proposition 1 summarizes two necessary conditions for rv-implementation. A first necessary condition is that only social choice functions which assign the same outcome to strategically indistinguishable payoff type profiles can be rv-implementable. We call such social choice functions dynamically robustly measurable, or briefly \(dr\)-measurable. A second necessary condition is \textit{ex-post incentive compatibility} (epIC). epIC requires that in the direct mechanism, truth-telling is a best response for every payoff type of an agent that expects his opponents to tell the truth, regardless of his belief about his opponents’ payoff types. epIC has been shown to be necessary for robust implementation in static mechanisms by Bergemann and Morris (2005), and is a strong incentive compatibility condition.\(^1\)

Proposition 2 provides sufficient conditions for rv-implementation. We call a social choice function \textit{strongly} \(dr\)-measurable with respect to some mechanism if for all agents, the social choice functions treats any two payoff types the same if \textit{that} mechanism does not strategically distinguish them. Proposition 2 says that under an economic assumption, for any mechanism in a large class of mechanisms, any epIC social choice function that is strongly \(dr\)-measurable with respect to the mechanism is rv-implementable. The proof of proposition 2 builds on the sufficiency proofs in Abreu and Matsushima (1992a,b).

From our results on strategic distinguishability (proposition 3) we know that in private consumption environments, \(dr\)-measurability, and strong \(dr\)-measurability with respect to the mechanism constructed in proposition 3 are weak conditions. Indeed, for generic valuation functions they are satisfied by \textit{all} social choice functions, and dynamic mechanisms can rv-implement a social choice functions if and only if it is epIC (corollary 1 in section 5).

1.1.3 A Specific Private Consumption Environment

Under RCIBR, dynamic mechanisms are just as powerful as static mechanisms, both in terms of strategically distinguishability and rv-implementation. Private consumption environments demonstrate that under RCSBR, dynamic mechanisms are much more powerful than static mechanisms. Example 1.1 highlights this by summarizing results in a private consumption

\(^1\)On the strength of epIC, see Jehiel, Meyer-Ter-Vehn, Moldovanu, and Zame (2006), but also see Bikhchandani (2006).
environment with specific valuation functions. Notably, the example points out that under RCSBR, dynamic mechanism can robustly virtually allocate a single object in an efficient manner in many cases the degree of preference interdependence prevents static mechanisms to do so.

Example 1.1 (compare BM, Section 3) An object is to be allocated to one of finitely many agents. Each agent $i$ has a finite payoff type space $\Theta_i$, $\{0,1\} \subseteq \Theta_i \subseteq [0,1]$, and receives utility $v_i(\theta_i, \theta_{-i})q_i + t_i$ if the payoff type profile is $(\theta_i, \theta_{-i})$, where $q_i$ is the probability that $i$ will receive the good, $t_i$ a monetary transfer and $v_i(\theta_i, \theta_{-i}) = \theta_i + \gamma \sum_{j \neq i} \theta_j$ the value of the object to $i$. The parameter $\gamma \geq 0$ measures the degree of preference interdependence in the environment. If $\gamma < \frac{1}{I-1}$ static mechanisms can strategically distinguish all payoff type profiles (see BM, Section 3). But if $\gamma \geq \frac{1}{I-1}$ neither static mechanisms nor dynamic mechanisms under RCIBR can strategically distinguish any payoff type profiles. In contrast, proposition 3 implies that for all $\gamma < \frac{1}{I-1}$ and for almost all $\gamma \geq \frac{1}{I-1}$, all payoff type profiles are strategically distinguishable by dynamic mechanisms if there is RCSBR.

If $\gamma \geq \frac{1}{I-1}$, only constant social choice functions are rv-implementable by static mechanisms or by dynamic mechanisms under RCIBR. This is because only constant social choice functions are robustly measurable (BM): only constant social choice functions assign the same outcome to payoff type profiles that are strategically indistinguishable static mechanisms, or, equivalently, strategically indistinguishable by dynamic mechanisms under RCIBR. In contrast, under RCSBR, all epIC social choice functions can be rv-implemented in dynamic mechanisms for almost all $\gamma$ by proposition 2. This includes in particular the efficient allocation of the object (a non-constant social choice function), which is epIC if the single-crossing condition $\gamma < 1$ holds.†

1.2 Related Literature

Robust Virtual Implementation. Carrying out the Wilson doctrine (Wilson, 1987) in the context of mechanism design, Bergemann and Morris (2005, 2009a,b), Chung and Ely (2007) and others determine which social choice functions are implementable if agents’ beliefs about others’ private information are not commonly known. This paper belongs to a subset of this literature on robust mechanism design that requires full (and not just partial) robust implementation, but slightly weakens the implementation concept to robust approximate, or robust virtual implementation. The present paper is more general than BM and Artemov, Kumimoto, and Serrano (2009), the other papers in this subset, inasmuch as we admit dynamic

---

†Let $I$ denote the number of agents. More precisely, an allocation rule is a function $q : \Theta \to \{0,1\}^I$ such that $\sum_{i=1}^I q_i(\theta) = 1$ for all $\theta \in \Theta$. It is efficient if for each payoff type profile $\theta$ and each agent $i$, $q_i(\theta) > 0$ implies $v_i(\theta) = \max_{j \neq i} v_j(\theta)$. Any efficient allocation rule is epIC for $\gamma < 1$ if combined with generalized VCG transfers, that is, if $i$ pays $q_i(\theta)(\gamma \sum_{j \neq i} \theta_j + \max_{j \neq i} \theta_j)$ (Dasgupta and Maskin, 2000).
mechanisms and not just static ones. At the same time, the present paper is less general inasmuch as we restrict attention to private consumption environments for the purpose of determining strategic distinguishability. From Artemov, Kunimoto, and Serrano (2009), we also differ in an additional dimension. Like BM, the present paper imposes no common knowledge assumptions on the (initial) belief hierarchies over payoff type profiles, while Artemov, Kunimoto, and Serrano adopt an intermediately robust approach. They assume that a finite set of first-order beliefs about payoff type profiles is common knowledge, and proceed to characterize (intermediate) robust virtual implementation.

**Dynamic Mechanisms.** Bergemann and Morris (2007) consider a complete information environment, and present an ascending price auction that robustly virtually allocates an object in an efficient manner even if there is so much preference interdependence that static mechanisms cannot. Because there is common knowledge of the payoff type profile among the bidders, Bergemann and Morris (2007) focus on robustness solely in terms of the uncertainty about others strategies. They use backward induction as their solution concept.

Penta (2009) explores robust implementation in incomplete information environments with multiple stages, where in each stage, an agent learns part of his payoff type and participates in playing a static mechanism. Penta’s work relates to ours as an agent can learn all of his payoff type at the first stage, and then participate in a sequence of static mechanisms, that is, participate in a dynamic mechanism (with only observable actions). He introduces the solution concept of backward rationalizability, which incorporates a belief revision assumption as expressed by common belief in future rationality at each decision node.

**Local Robustness.** While in this paper, robustness means “global robustness” (mechanisms should work regardless of agents’ initial beliefs about others’ payoff types), robustness sometimes also refers to a “local robustness” notion (mechanisms should work in a neighborhood of some belief hierarchy). Game theoretic results show that already small perturbations of higher order beliefs can change the set of rationalizable strategies and the set of Bayesian Nash equilibria of a mechanism (see Rubinstein, 1989; Weinstein and Yildiz, 2007).† As a mechanism design counterpart to these results Oury and Tercieux (2009) prove that requiring local robustness of partial (Bayesian) implementation is equivalent to requiring full rationalizable implementation; Di Tillio (2009) proves that full rationalizable implementation is locally robust.

†Even if the mechanism designer is certain of the exact infinite belief hierarchies over payoff type profiles the set of Bayesian Nash equilibria can depend on the type space giving rise to the belief hierarchies. A type space giving rise to the same belief hierarchies as the naïve type space can yield a different set of Bayesian Nash equilibria (see Ely and Pęski, 2006; Dekel, Fudenberg, and Morris, 2007; Sadzik, 2009).
2 Example

Example 2.1 argues in a simple environment that static mechanisms or, equivalently, dynamic mechanisms under RCIBR can strategic distinguish less payoff type profiles than dynamic mechanisms under RCSBR (assuming belief-completeness).

**Example 2.1** There are two agents, $i \in \{1, 2\}$, with two conceivable payoff types each, $\hat{\theta}_i \in \{\theta_i, \theta'_i\}$, and four outcomes, $w, x, y, z$. The utility functions are given in figure 1. We will show that all payoff type profiles are a) strategically indistinguishable by static mechanisms (compare BM, proposition 1), b) strategically indistinguishable by dynamic mechanisms if there is only RCIBR (compare Müller, 2009), but c) strategically distinguishable by dynamic mechanisms if there is RCSBR.

a) Let $S_1$ and $S_2$ be finite sets and $\Gamma : S_1 \times S_2 \rightarrow \{w, x, y, z\}$ be a static mechanism. Let $s^1_1 \in S_1$ be a best response for payoff type $\theta_1$ who is certain that agent 2’s payoff type is $\theta_2$ and that agent 2 will play some arbitrarily chosen $s^0_2 \in S_2$. Since $u_1(\cdot, \theta_1, \theta_2) = u_1(\cdot, \theta'_1, \theta_2)$, $s^1_1$ must then also be a best response for payoff type $\theta'_1$ who is certain that agent 2’s payoff type is $\theta'_2$ and that agent 2 will play $s^0_2 \in S_2$. Hence, $s^1_1$ survives one round of iterated elimination of never-best responses for both payoff types of agent 1 — $s^1_1$ is rational for both payoff types of agent 1. Since $u_2(\cdot, \theta_1, \theta'_2) = u_2(\cdot, \theta'_1, \theta_2)$ there also is a $s^1_2 \in S_2$ that is rational for both payoff types of agent 2. We can now iterate this argument: Let $s^2_1$ be a best response for $\theta_1$ to the degenerate belief in $(s^1_2, \theta_2)$. Then $s^2_1$ is a best response for $\theta'_1$ to the degenerate belief in $(s^0_2, \theta_2)$. $s^2_1$ and an analogously derived $s^2_2 \in S_2$ survive two rounds of iterated elimination of never-best responses for both payoff types of the respective agent. For any $(\hat{\theta}_1, \hat{\theta}_2)$, $(s^2_1, s^2_2, \hat{\theta}_1, \hat{\theta}_2)$ is consistent with rationality and mutual belief in rationality. There will be a $k \in \mathbb{N}$ at which the iterated elimination procedure stops. $(s^k_1, s^k_2, \hat{\theta}_1, \hat{\theta}_2)$ is consistent with rationality and common belief in rationality. The strategy profile $(s^k_1, s^k_2)$ is interim correlated rationalizable for every payoff type profile. Therefore, no inference about the agents’ payoff types can be drawn from $(s^k_1, s^k_2)$. Formally, $(s^k_1, s^k_2)$ is a terminal node that can be strongly rationalizably reached by all payoff type profiles. Hence, all payoff type profiles are strategically indistinguishable. This does not change if we admit lotteries over $\{w, x, y, z\}$ as outcomes of mechanisms.

<table>
<thead>
<tr>
<th>$u_1(\cdot, \theta_1, \theta_2)$</th>
<th>$w$</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1(\cdot, \theta_1, \theta'_2)$</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>$u_1(\cdot, \theta'_1, \theta_2)$</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$u_1(\cdot, \theta'_1, \theta'_2)$</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>$u_1(\cdot, \theta'_1, \theta'_2)$</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$u_2(\cdot, \theta_1, \theta_2)$</th>
<th>$w$</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_2(\cdot, \theta_1, \theta'_2)$</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$u_2(\cdot, \theta_1, \theta'_2)$</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$u_2(\cdot, \theta'_1, \theta_2)$</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>$u_2(\cdot, \theta'_1, \theta'_2)$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1: Utility functions
b) The argument is essentially as in a): Take any dynamic mechanism, and let $s_1^1$ be a sequential best response for payoff type $\theta_1$ whose beliefs are as follows: initially, $\theta_1$ is certain that agent 2’s payoff type is $\theta_2$ and that agent 2 will play some arbitrarily chosen $s_2^0$. If surprised, $\theta_1$ continues to be certain that agent 2’s payoff type is $\theta_2$ and believes that agent 2 plays some arbitrarily chosen strategy that admits the current information set. Then, $s_1^1$ must also be a sequential best response for payoff type $\theta_1'$ who at each information set is certain that agent 2’s payoff type is $\theta_2'$ and that agent 2 plays the strategy that $\theta_1$ believes in. Since there is also a $s_2^1$ that is a sequential best response for both $\theta_2$ and $\theta_2'$, we can again iterate the argument to find a strategy profile $(s_1^k, s_2^k)$ that is consistent with RCIBR for every payoff type profile.

c) If there is RCSBR, the mechanism presented in figure 2 strategically distinguishes all payoff type profiles. To see this, observe that

- it is never rational for $\theta_2$ to play $\vartheta_2'$ if agent 1 announced $\vartheta_1'$ and
  it is never rational for $\theta_2'$ to play $\vartheta_2$ if agent 1 announced $\vartheta_1$, therefore

- if agent 1 (strongly) believes in agent 2’s rationality, $\vartheta_1'$ is never rational for $\theta_1$, therefore

- if agent 2 strongly believes in agent 1’s rationality and 1’s (strong) belief in 2’s rationality, and if agent 1 announced $\vartheta_1'$, then agent 2 concludes that 1’s payoff type is $\theta_1'$ — RCBSR allows us to predict agent 2’s belief about agent 1’s payoff type — and $\vartheta_2$ is never rational for $\theta_2'$, therefore

- if agent 1 (strongly) believes in ..., $\vartheta_1$ is never rational for $\theta_1'$, therefore

- if agent 2 strongly believes in ..., and if agent 1 announced $\vartheta_1$, then agent 2 concludes that 1’s payoff type is $\theta_1$ — again, RCBSR allows us to predict agent 2’s belief about agent 1’s payoff type — and $\vartheta_2'$ is never rational for $\theta_2$.

![Figure 2: Mechanism that strategically distinguishes all payoff type profiles](image)

That is, truth-telling is the unique strongly rationalizable strategy for both payoff types of both agents. No assumptions about the agents’ initial beliefs about each others’ payoff types are necessary. Agent 2 “learns” agent 1’s payoff type during the course of the mechanism, and
for any fixed belief about agent 1’s payoff type, 2’s payoff types have different preferences. Note that if nothing were known about how agent 2 revises his beliefs, the above chain of implications would break down because we could never exclude the case that agent 2, once surprised by seeing $\vartheta'_1$, believes to face payoff type $\theta_1$.

The mechanism of figure 2 robustly implements the non-constant epIC social choice function $f$ defined by $f(\theta_1, \theta_2) = w, f(\theta_1, \theta'_2) = x, f(\theta'_1, \theta_2) = y$ and $f(\theta'_1, \theta'_2) = z$. Note that no static mechanism can robustly, or even robustly virtually implement $f$ (all payoff type profiles are strategically indistinguishable by static mechanisms, hence merely constant social choice functions rv-implementable by static mechanisms). Proposition 2 will provide a dynamic mechanism that rv-implements any epIC social choice function in this example if outcomes are lotteries over $\{w, x, y, z\}$.

3 Environment and Preliminaries

There is a finite set $I = \{1, \ldots, I\}$ of agents\(^1\); we assume $I \geq 2$. Each agent $i \in I$ has a finite payoff type space $\Theta_i$. There is a finite set $X$ of pure outcomes; the set of outcomes is the set $Y = \Delta(X)$ of lotteries (that is, probability measures) over $X$. Agent $i \in I$ has a von Neumann-Morgenstern utility function $u_i : X \times \Theta \to \mathbb{R}$. Abusing notation slightly, we let $u_i$ also denote the (expected) utility function that maps $(y, \theta) \in \mathbb{R}^{\#X} \times \Theta$ to $u_i(y, \theta) = \sum_{x \in X} y(x) \cdot u_i(x, \theta)$. Let $\bar{y}$ denote the uniform lottery that assigns probability $\frac{1}{\#X}$ to all $x \in X$.

For any family $(Z_i)_{i \in I}$ of sets $Z_i$, $Z$ denotes the Cartesian product $\prod_{i \in I} Z_i$.\(^2\) It is also understood that $z$ denotes $(z_1, \ldots, z_I)$ whenever $z_i \in Z_i$ for all $i \in I$. If $Z_i = A_i \times B_i$ for all $i \in I$, we sometimes ignore the correct order of tuples and write $((a_1, \ldots, a_I), (b_1, \ldots, b_I)) \in Z$ for $(a_i, b_i)_{i \in I} \in Z$. If $m > n$, $m, n \in \mathbb{N} = \{0, 1, \ldots\}$, then $\{m, \ldots, n\}$ denotes the empty set.

3.1 Mechanisms

In this subsection we recall the definition of an extensive game form (see e.g. Kuhn (1953)). The class of mechanisms we consider is the class of finite extensive game forms with perfect recall and only non-trivial decision nodes. Perfect recall and the exclusion of trivial decision nodes ensure that our definition of a Bayesian agent (made in subsection 3.2) is sensible.

Definition 1 An extensive game form is a tuple $\Gamma = \langle H, (\mathcal{H}_i)_{i \in I}, P, C \rangle$ such that

- $H$ is a nonempty finite set of finite sequences with codomain $A$ (where $A$ is a nonempty

\(^1\)Note we let $I$ denote both the set of agents and its cardinality.

\(^2\)As an exception to this rule, $H_i$ will denote the set of non-terminal histories at which $i$ is active, and $H = \prod_{i \in I} H_i$ the set of all histories (compare definition 1).
set of actions) such that with $h$ every initial subsequence of $h$ is in $H$.\footnote{Let $A$ be a nonempty set. A finite sequence $h$ of length $n \in \mathbb{N}$ with codomain $A$ is a function $h : \{1, \ldots, n\} \to A$. A finite sequence $g : \{1, \ldots, k\} \to A$ is an initial subsequence of the finite sequence $h : \{1, \ldots, n\} \to A$ if $k \leq n$ and $g = h_l$ for all $l \in \{1, \ldots, k\}$. Note that $\emptyset$ (the unique finite sequence mapping $\{1, \ldots, 0\}$ to $A$) is an initial subsequence of every finite sequence with codomain $A$. For $h : \{1, \ldots, n\} \to A$ and $a \in A$, $(h, a)$ denotes the finite sequence that maps $\{1, \ldots, n + 1\}$ into $A$, has $h$ as an initial subsequence and maps $n + 1$ to $a$.} We let $A(h) = \{a \in A ; (h, a) \in H\}$ for $h \in H$, $T = \{h \in H; A(h) = \emptyset\}$ and call $\emptyset \in H$ the initial history. We write $h' \preceq h$ if $h' \in H$ is an initial subsequence of $h \in H$, and $h' < h$ if $h' \preceq h$ and $h' \neq h$.

- $P : H \setminus T \to I$. We let $H_i = \{h \in H \setminus T ; P(h) = i\}$ for all $i \in I$.\footnote{For notational convenience, we require $H_i \neq \emptyset$. This ensures that $i$'s strategy set is nonempty (and thus also excludes trivial mechanisms with $H = \{\emptyset\}$).}

- for each $i \in I$, $\mathcal{H}_i$ is a partition\footnote{A partition of $H_i$ is a family $(\mathcal{H}_n)_{n=1}^N$ of nonempty, pairwise disjoint sets $\mathcal{H}_n \subseteq H_i$ such that $\bigcup_{n=1}^N \mathcal{H}_n = H_i$.} of $H_i$ such that
  - for all $\mathcal{H} \in \mathcal{H}_i$ and all $h, h' \in \mathcal{H}$, $A(h) = A(h')$. For $h \in H_i$ we let $[h]$ denote the element of $\mathcal{H}_i$ containing $h$, and $A([h])$ denote $A(h)$.
  - for all $\mathcal{H} \in \mathcal{H}_i$ and all $h, h' \in H$, if $h \in \mathcal{H} \in \mathcal{H}_i$ and $h' < h$ then $h' \notin \mathcal{H}$.

- $C : T \to Y$.

$H$ is interpreted as set of histories $h$. A history $h = (a_1, \ldots, a_n)$ is a finite sequence of actions and can be either terminal ($h \in T$) or non-terminal ($h \in H \setminus T$). The game starts at the initial history and ends once a terminal history is reached. $P(h)$ is the agent or player who is active at the non-terminal history $h$. At $h$, player $P(h)$ only knows he is at one of the histories in the information set $[h]$, and he can choose an action from $A([h])$. History $h = (a_1, \ldots, a_n)$ obtains if $P(\emptyset)$ chooses $a_1$ at the initial history, $P(a_1)$ chooses $a_2$ at history $(a_1)$, ..., and $P(a_1, \ldots, a_{n-1})$ chooses $a_n$ at history $(a_1, \ldots, a_{n-1})$. The outcome function $C$ assigns an outcome to each terminal history. Note that we do neither allow for infinitely many time periods (histories are finite sequences) nor for infinitely many actions at any history ($H$ is finite). Also note that $\preceq$ partially orders $H$.

A strategy for player $i$ in an extensive game form $\Gamma$ is a function $s_i : \mathcal{H}_i \to A$ such that $s_i(\mathcal{H}) \in A(\mathcal{H})$ for all $\mathcal{H} \in \mathcal{H}_i$. We let $S_i$ be the set of $i$'s strategies. $\zeta(s) \in H$ denotes the terminal history induced by strategy profile $s = (s_i)_{i \in I} \in S$. We let $l_h$ denote the length of history $h \in H$. For $t \in \{0, \ldots, \max_{h \in H} l_h\}$, $H^{=t} = \{h \in H ; l_h = t\}$ denotes the set of histories with length $t$. For $h' \in H$, $H^{\geq h'} = \{h \in H ; h \succeq h'\}$ is the set of histories that follow $h'$. $H^{\leq t}$, $H^{\geq h'}$, $H_{\overset{\leq t}}^{h'}$, etc. are defined analogously. For each player $i \in I$, each history $h \in H$ and each information set $\mathcal{H} \in \mathcal{H}_j$, $j \in I$,

$$S_i(h) = \{s_i \in S_i ; s_i \text{ admits } h\} = \{s_i \in S_i ; \exists s_{-i} \in S_{-i} ; h \leq \zeta(s)\},$$

\footnote{A partition of $H_i$ is a family $(\mathcal{H}_n)_{n=1}^N$ of nonempty, pairwise disjoint sets $\mathcal{H}_n \subseteq H_i$ such that $\bigcup_{n=1}^N \mathcal{H}_n = H_i$.}
\[ S_i(\mathcal{H}) = \{ s_i \in S_i; s_i \text{ admits } \mathcal{H} \} = \{ s_i \in S_i; \exists s_{-i} \in S_{-i} \exists h \in \mathcal{H} : h \preceq \zeta(s) \}; \]

\[ S_{-i}(\mathcal{H}) \text{ etc. are defined similarly.} \]

We let \( \Sigma_i = S_i \times \Theta_i \), \( \Sigma_i(h) = S_i(h) \times \Theta_i \) be the set of possible strategy-payoff type pairs not preventing \( h \in H \) and \( \Sigma_{-i}(\mathcal{H}) = S_{-i}(\mathcal{H}) \times \Theta_{-i} \) etc. Moreover, for any \( i \in I \), \( J \subseteq I \), \( (s_j)_{j \in J} \in \prod_{j \in J} S_j \) and any \( \theta \in \Theta \),

\[ H((s_j, \theta_j)_{j \in J}) = H((s_j)_{j \in J}) = \{ h \in H; (s_j)_{j \in J} \text{ admits } h \} = \{ h \in H; \exists (s_j)_{j \in I \setminus J} \in \prod_{j \in I \setminus J} S_j : h \preceq \zeta(s) \}, \]

\[ \mathcal{H}((s_j, \theta_j)_{j \in J}) = \mathcal{H}((s_j)_{j \in J}) = \{ \mathcal{H} \in \mathcal{H}_i; \exists h \in \mathcal{H} : h \in H((s_j)_{j \in J}) \}. \]

In particular, for any strategy profile \( s \in S \), \( H(s) \) consists of all histories on the path induced by \( s \). Combinations with previously defined notation for sets of histories have the obvious meaning, e.g. \( H^<\zeta_{\mathcal{H}}(s_i) = H_i \cap H(s_i) \cap H^<\zeta_{\mathcal{H}} \), and \( H^=\zeta_{\mathcal{H}}(s) \) is the singleton consisting of the initial subsequence of \( \zeta(s) \) with length 1. For \( A \subseteq \Sigma_i \), \( H(A) = \bigcup_{(s_i, \theta_i) \in A} H(s_i, \theta_i) \) etc.

**Definition 2** A (dynamic) mechanism is an extensive game form \( \Gamma = (H, (\mathcal{H}_i)_{i \in I}, P, C) \) such that

- (perfect recall) for all \( i \in I \), \( s_i \in S_i \) and \( \mathcal{H} \in \mathcal{H}_i \), if \( \mathcal{H} \cap H(s_i) \neq \emptyset \) then \( \mathcal{H} \subseteq H(s_i) \).

- (no trivial decisions) for all \( (h, a) \in H \) there exists an action \( a' \neq a \) such that \( (h, a') \in H \).

We define a binary relation \( \preceq \) on \( \mathcal{H}_i \) by \( \mathcal{H}' \preceq \mathcal{H} \) if there are \( h' \in \mathcal{H}' \) and \( h \in \mathcal{H} \) such that \( h' \preceq h \). We extend \( \preceq \) to \( \mathcal{H}_i \subseteq \mathcal{H}_i \cup \{ \emptyset \} \) (if necessary) by letting \( \{ \emptyset \} \preceq \mathcal{H} \) for all \( \mathcal{H} \in \mathcal{H}_i \). \( \mathcal{H}_i \preceq \mathcal{H} \) etc. have the obvious meaning.

### 3.2 Beliefs and Sequential Rationality

Player \( i \)'s beliefs are captured by a family of probability measures on \( \Sigma_{-i} \), with each measure representing \( i \)'s belief at one of his information sets or at the initial history, so at one of the elements of \( \mathcal{H}_i \). The measures representing \( i \)'s beliefs are summarized by a conditional probability system.

**Definition 3** (Myerson, 1986) Let \( i \in I \). A conditional probability system (CPS) on \( \Sigma_{-i} \) is a function \( \mu_i : 2^{\Sigma_{-i}} \times \mathcal{H}_i \to [0, 1] \) such that

a) for all \( \mathcal{H} \in \mathcal{H}_i \), \( \mu_i(\cdot | \mathcal{H}) \) is a probability measure on \((\Sigma_{-i}, 2^{\Sigma_{-i}})\)

\(1\)Note that for any history \( h \), the Cartesian product of the sets \( S_i(h), \ldots, S_i(h) \) equals the set of strategy profiles admitting \( h \), \( S(h) = \{ s \in S; h \preceq \zeta(s) \} \). For any information set \( \mathcal{H} \), the Cartesian product \( \prod_{i \in I} S_i(\mathcal{H}) \) is a superset (but not necessarily a subset) of the set \( S(\mathcal{H}) = \{ s \in S; \exists h \in \mathcal{H}; h \preceq \zeta(s) \} \).
b) for all \( \mathcal{H} \in \mathcal{H}_i \), \( \mu_i(\Sigma_{\neg i}(\mathcal{H})|\mathcal{H}) = 1 \).

c) for all \( \mathcal{H}, \mathcal{H}' \in \mathcal{H}_i \), if \( \mathcal{H}' \preceq \mathcal{H} \) then \( \mu_i(A|\mathcal{H})\mu_i(\Sigma_{\neg i}(\mathcal{H})|\mathcal{H}') = \mu_i(A|\mathcal{H}') \) for all \( A \in 2^{\Sigma_{\neg i}} \).

Condition b) requires that at information set \( \mathcal{H} \) agent \( i \) cannot put strictly positive (marginal) probability on any strategy of \( \neg i \) which would have prevented that \( \mathcal{H} \) occurs. Condition c) demands that \( i \) uses Bayesian updating “whenever applicable”: Suppose \( \mathcal{H}' \preceq \mathcal{H} \), and that at \( \mathcal{H}' \), \( i \) estimates that \( A \) is going to happen with probability \( \mu_i(A|\mathcal{H}') \). The play proceeds and \( i \) finds himself at \( \mathcal{H} \). If \( \mathcal{H} \) was no surprise to him (that is, if \( \mu_i(\Sigma_{\neg i}(\mathcal{H})|\mathcal{H}') > 0 \)) he should now believe in \( A \) with probability

\[
\mu_i(A|\mathcal{H}) = \frac{\mu_i(A|\mathcal{H}')}{\mu_i(\Sigma_{\neg i}(\mathcal{H})|\mathcal{H}')},
\]

but if \( \mathcal{H} \) did surprise him (that is, if \( \mu_i(\Sigma_{\neg i}(\mathcal{H})|\mathcal{H}') = 0 \) ) condition c) allows any new estimate of the likelihood of \( A \), i.e. \( \mu_i(A|\mathcal{H}) \in [0, 1] \).

We let \( \Delta(\Sigma_{\neg i}) \) denote the set of probability measures on \( \Sigma_{\neg i} \) and \( \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \) denote the set of conditional probability systems on \( \Sigma_{\neg i} \). Given \( \mu_i \in \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \) let

\[
U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}) = \int_{\Sigma_{\neg i}(\mathcal{H})} u_i(C(\zeta(s)), \theta)\mu_i(d(s_{\neg i}, \theta_{\neg i})|\mathcal{H})
\]

define \( U_i^{\mu_i} : \{(s_i, \theta_i, \mathcal{H}) \in \Sigma_i \times \mathcal{H}_i; s_i \in S_i(\mathcal{H})\} \to \mathbb{R} \). \( U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}) \) is player \( i \)'s expected utility if he plays \( s_i \), is of payoff type \( \theta_i \) and holds beliefs \( \mu_i(\cdot|\mathcal{H}) \).

**Definition 4** Strategy \( s_i \in S_i \) is sequentially rational for payoff type \( \theta_i \in \Theta_i \) of player \( i \) with respect to beliefs \( \mu_i \in \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \) if for all \( \mathcal{H} \in \mathcal{H}_i(s_i) \) and all \( s'_i \in S_i(\mathcal{H}) \)

\[
U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}) \geq U_i^{\mu_i}(s'_i, \theta_i, \mathcal{H}).
\]

We let \( r_i : \Theta_i \times \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \to S_i \) denote the correspondence that maps \( (\theta_i, \mu_i) \) to the set of strategies that are sequentially rational for payoff type \( \theta_i \) with beliefs \( \mu_i \), and \( \rho_i : \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \to \Sigma_i \) denote the correspondence that maps \( \mu_i \) to the subset of \( \Sigma_i \) consisting of strategy-payoff type pairs \( (s_i, \theta_i) \) such that \( s_i \) is sequentially rational for payoff type \( \theta_i \) with beliefs \( \mu_i \). For each \( i \in I \), \( r_i \) and \( \rho_i \) are nonempty-valued.

### 3.3 Strong Rationalizability

Battigalli (2003) defines strong rationalizability for multi-stage games. We extend his definition to dynamic mechanisms.

**Definition 5** For \( i \in I \) let \( F_i^0 = \Sigma_i \) and \( \Phi_i^0 = \Delta_{\mathcal{H}_i}(\Sigma_{\neg i}) \) and recursively define the set \( F_i^{k+1} \) of strongly \( k \)-rationalizable pairs \( (s_i, \theta_i) \) for player \( i \) by

\[
F_i^{k+1} = \rho_i(\Phi_i^k),
\]

13
and the set $\Phi_{i}^{k+1}$ of strongly $k$-rationalizable beliefs for player $i$ by

$$
\Phi_{i}^{k+1} = \left\{ \mu_{i} \in \Phi_{i}^{k}; \forall H \in \mathcal{H}_{i} \left( \Sigma_{-i}(H) \cap F_{i}^{k+1} \neq \emptyset \Rightarrow \mu_{i}(F_{i}^{k+1}|H) = 1 \right) \right\},
$$

$k \in \mathbb{N}$. Finally, let $F_{i}^{\infty} = \bigcap_{k=0}^{\infty} F_{i}^{k}$ be the set of strongly rationalizable strategy-payoff type pairs for player $i$, and $\Phi_{i}^{\infty} = \bigcap_{k=0}^{\infty} \Phi_{i}^{k}$ be the set of strongly rationalizable beliefs for player $i$.

The strongly rationalizable strategies are determined by iteratively deleting never-best sequential responses, where it is required that at each of his information sets an agent believes in the highest degree of his opponents’ rationality that is consistent with the information set (best-rationalization principle). For convenience, we let $R_{i}^{k}(\theta_{i}) = \{ s_{i} \in S_{i}; (s_{i}, \theta_{i}) \in F_{i}^{k} \}$ denote the set of strongly (k-)rationalizable strategies for $\theta_{i} \in \Theta_{i}$, where $k \in \mathbb{N} \cup \{\infty\}$ and $i \in I$. The sets $R_{i}^{\infty}(\theta_{i})$ and $\Phi_{i}^{\infty}$ are nonempty for all $i \in I$ and $\theta_{i} \in \Theta_{i}$.

### 4 Robust Virtual Implementation

A social choice function (scf) is a function $f : \Theta \rightarrow Y$. It assigns a desired outcome to each payoff type profile. A social choice function is robustly implementable if there exists a mechanism in which, for every payoff type profile $\theta$, every strongly rationalizable strategy profile leads to $f(\theta)$. A social choice functions is robustly virtually implementable if it can be robustly approximately implemented in the following sense.

**Definition 6** Social choice function $f$ is robustly $\varepsilon$-implementable for $\varepsilon > 0$ if there is a mechanism $\Gamma$ such that $\|C(\zeta(s)) - f(\theta)\| \leq \varepsilon$ for all $(s, \theta) \in F_{i}^{\infty}$.

A second necessary condition for robust and robust virtual implementation is that the social choice function treats strategically indistinguishable payoff type profiles the same. We write $\theta \sim^{\Gamma} \theta'$ and say that the payoff type profiles $\theta \in \Theta$ and $\theta' \in \Theta$ are $\Gamma$-strategically indistinguishable.

**Definition 7** Social choice function $f$ is ex-post incentive compatible (epIC) if for all $i \in I$, all $\theta \in \Theta$ and all $\theta'_{i} \in \Theta_{i}$

$$
u_{i}(f(\theta), \theta) \geq u_{i}(f(\theta'_{i}, \theta_{-i}), \theta).
$$

As shown by Bergemann and Morris (2005), ex-post incentive compatibility is necessary for robust implementation. Admitting dynamic mechanisms and being content with robust virtual implementation do not change this.

**4.1 Necessary Conditions for Robust Virtual Implementation**

A second necessary condition for robust and robust virtual implementation is that the social choice function treats strategically indistinguishable payoff type profiles the same. We write $\theta \sim^{\Gamma} \theta'$ and say that the payoff type profiles $\theta \in \Theta$ and $\theta' \in \Theta$ are $\Gamma$-strategically indistinguishable.
indistinguishable if \( \Gamma \) is a mechanism and \( \zeta(s) = \zeta(s') \) for some \( s \in R^{\mathcal{I},\infty}(\theta), s' \in R^{\mathcal{I},\infty}(\theta') \). We write \( \theta \sim \theta' \) and say \( \theta \) and \( \theta' \) are **strategically indistinguishable** if \( \theta \sim \Gamma \theta' \) for every mechanism \( \Gamma \). The binary relations \( \sim \) and \( \sim^\Gamma \) are reflexive and symmetric, but not necessarily transitive.

**Definition 8** \( \text{Scf} \) is dynamically robustly measurable (dr-measurable) if for all \( \theta, \theta' \in \Theta \), \( \theta \sim \theta' \) implies \( f(\theta) = f(\theta') \).

**Proposition 1** If \( \text{scf} \ f \) is rv-implementable, then \( f \) is epIC and dr-measurable.

**Proof.** We first show \( f \) is dr-measurable. Suppose \( \theta \sim \theta' \). Take \( \varepsilon > 0 \), then there is a mechanism \( \Gamma \) that robustly \( \varepsilon \)-implements \( f \). Since \( \theta \sim \theta' \), there are \( s \in R^\infty(\theta) \) and \( s' \in R^\infty(\theta') \) such that \( \zeta(s) = \zeta(s') \). By robust \( \varepsilon \)-implementation, \( \|C(\zeta(s)) - f(\theta)\| \leq \varepsilon \) and \( \|C(\zeta(s')) - f(\theta')\| \leq \varepsilon \) and thus \( \|f(\theta) - f(\theta')\| \leq 2\varepsilon \). Since this holds for all \( \varepsilon > 0 \), \( f(\theta) = f(\theta') \).

Now we establish by a direct proof that \( f \) is epIC. Suppose \( f \) is robustly virtually implementable, and take any \( i \in I, \theta_i, \theta'_i \in \Theta_i \) and \( \theta_{-i} \in \Theta_{-i} \). We are going to show that

\[
\begin{align*}
    u_i(f(\theta), \theta) &\geq u_i(f(\theta', \theta_{-i}), \theta).
\end{align*}
\]

If \( f(\theta'_i, \theta_{-i}) = f(\theta) \) then (1) is trivially satisfied, thus consider the case where \( f(\theta'_i, \theta_{-i}) \neq f(\theta) \).

Let \( \varepsilon \) satisfy \( 0 < \varepsilon < \frac{1}{2}\|f(\theta'_i, \theta_{-i}) - f(\theta)\| \), then there is a mechanism \( \Gamma = (H, (\mathcal{H}_i)_{i \in I}, P, C) \) that robustly \( \varepsilon \)-implements \( f \), that is, a mechanism \( \Gamma \) such that \( \|C(\zeta(s)) - f(\tilde{\theta})\| \leq \varepsilon \) for all \( (\tilde{s}, \tilde{\theta}) \in F^{\infty} \).

For each \( j \neq i \), pick some \( s_j \in R^\infty_j(\theta_j) \). Let \( \lambda_i \in \Delta(\Sigma_{-i}) \) denote the point belief in \( (s_{-i}, \theta_{-i}) \), and let \( \mu'_i \) be an element of \( \Phi^\infty_i \). Define \( \mu_i : 2^{\Sigma_{-i}} \times \mathcal{H} \rightarrow [0, 1] \) by \( \mu_i(\cdot | H) = \lambda_i \) for \( H \in \mathcal{H}_i(s_{-i}) \) and \( \mu_i(\cdot | \mathcal{H}) = \mu'_i(\cdot | \mathcal{H}) \) for \( \mathcal{H} \notin \mathcal{H}_i(s_{-i}) \). Note that \( \mu_i \) is a CPS. Indeed, since \( \mu'_i \in \Phi^\infty_i \) and all the mass of \( \lambda_i \) concentrates on a profile of strongly rationalizable strategy-payoff type pairs, \( \mu_i \in \Phi^\infty_i \). Hence \( \rho_i(\mu_i) \subseteq F^\infty_i \) — if \( \tilde{s}_i \) is sequentially rational for \( \tilde{\theta}_i \) with respect to \( \mu_i \) then \( (\tilde{s}_i, \tilde{\theta}_i) \) is strongly rationalizable.

Pick some \( s_i \in r_i(\theta_i, \mu_i) \) and some \( s'_i \in r_i(\theta'_i, \mu_i) \). Since \( \Gamma \) robustly \( \varepsilon \)-implements \( f \)

\[
\begin{align*}
    \|C(\zeta(s'_i, s_{-i})) - C(\zeta(s))\| &\geq \|f(\theta_i', \theta_{-i}) - f(\theta_i)\| - 2\varepsilon > 0
\end{align*}
\]

and thus \( \zeta(s'_i, s_{-i}) \neq \zeta(s) \). So \( \mathcal{H}_i(s) \neq \emptyset \) (otherwise \( \zeta(s'_i, s_{-i}) = \zeta(s) \)). What is more, there is a (unique) information set \( \mathcal{H}' \in \mathcal{H}_i(s) \) such that \( s_i(\mathcal{H}') \neq s'_i(\mathcal{H}') \) and \( s_i(\mathcal{H}) = s'_i(\mathcal{H}) \) for all \( \mathcal{H} \in \mathcal{H} \). By the definition of sequential rationality, \( \forall \mathcal{H} \in \mathcal{H}_i(s_i) \forall \tilde{s}_i \in S_i(\mathcal{H}) : U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}) \geq U_i^{\mu_i}(\tilde{s}_i, \theta_i, \mathcal{H}) \).

In particular,

\[
\begin{align*}
    U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}') \geq U_i^{\mu_i}(s'_i, \theta_i, \mathcal{H}')
\end{align*}
\]
Since \( \mu_i(\cdot|\mathcal{H}') = \lambda_i \),
\[
U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}') = \int_{\Sigma_{-i}(\mathcal{H}')} u_i(C(\zeta(s)), \theta) \mu_i(d(s_{-i}, \theta_{-i})|\mathcal{H}')
\]
\[= u_i(C(\zeta(s)), \theta).\]

Since \( \Gamma \) robustly \( \varepsilon \)-implements \( f \), \( \|u_i(C(\zeta(s)), \theta) - u_i(f(\theta), \theta)\| \leq K \cdot \varepsilon \), where \( K \) denotes the Lipschitz constant of \( u_i(\cdot, \theta) \).\(^\dagger\) Similarly, \( \|U_i^{\mu_i}(s_i', \theta_i, \mathcal{H}') - u_i(f(\theta_i', \theta_{-i}), \theta)\| \leq K \cdot \varepsilon \), and so
\[
u_i(f(\theta), \theta) - u_i(f(\theta_i', \theta_{-i}), \theta)
\]
\[
\geq u_i(f(\theta), \theta) - U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}') + U_i^{\mu_i}(s_i', \theta_i, \mathcal{H}') - u_i(f(\theta_i', \theta_{-i}), \theta)
\]
\[
\geq -\|u_i(f(\theta), \theta) - U_i^{\mu_i}(s_i, \theta_i, \mathcal{H}')\| - \|U_i^{\mu_i}(s_i', \theta_i, \mathcal{H}') - u_i(f(\theta_i', \theta_{-i}), \theta)\|
\]
\[
\geq -2K \cdot \varepsilon.
\]

Since this holds for every sufficiently small \( \varepsilon > 0 \), (1) follows. \( \Box \)

### 4.2 Sufficient Conditions for Robust Virtual Implementation

In this subsection, we show that ex-post incentive compatibility and a strong version of dr-measurability are sufficient for robust virtual implementation. As in BM and Artemov, Kunitomo, and Serrano (2009), the implementing mechanism is constructed according to the ideas of Abreu and Matsushima (1992a,b). We assume the same economic property as BM.

**Definition 9 (Economic Property)** The economic property is satisfied if there exists a profile of lotteries \( (z_i)_{i \in I} \) such that for each \( i \in I \) and \( \theta \in \Theta \), both \( u_i(z_i, \theta) > u_i(\bar{y}, \theta) \) and \( u_j(\bar{y}, \theta) \geq u_j(z_i, \theta), j \neq i \).

The economic property is satisfied e.g. in quasilinear environments and in example 2.1 (if lotteries are admitted)\(^\dagger\).

Let \( i \in I \), \( \theta_i, \theta_i' \in \Theta_i \). We write \( \theta_i \sim_{\Gamma} \theta_i' \) and say that payoff types \( \theta_i \) and \( \theta_i' \) are \( \Gamma \)-strategically indistinguishable if \( \Gamma \) is a mechanism and there exists \( \theta_{-i} \in \Theta_{-i} \) such that \( \theta \sim_{\Gamma} (\theta_i', \theta_{-i}) \). We write \( \theta_i \sim \theta_i' \) and say that \( \theta_i \) and \( \theta_i' \) are *strategically indistinguishable* if \( \theta_i \sim_{\Gamma} \theta_i' \) for every mechanism \( \Gamma \).

**Definition 10** Social choice function \( f \) is

- *strongly \( \Gamma \)-measurable* if for all \( i \in I \) and \( \theta_i, \theta_i' \in \Theta_i \), \( \theta_i \sim_{\Gamma} \theta_i' \) implies \( f(\theta_i, \theta_{-i}) = f(\theta_i', \theta_{-i}) \) for all \( \theta_{-i} \in \Theta_{-i} \).

\(^\dagger\)Clearly \( u_i(\cdot, \theta) : \mathbb{R}^{N_i} \to \mathbb{R} \) is Lipschitz continuous: \( \|u_i(y, \theta) - u_i(y', \theta)\| = \|(y - y') \cdot (u_i(x, \theta))_{x \in X}\| \leq \|y - y'\| \cdot \|(u_i(x, \theta))_{x \in X}\| \).

\(^\dagger\)In example 2.1, \( z_1 = w \) and \( z_2 = z \).
• strongly dr-measurable if for all \( i \in I \) and \( \theta_i, \theta'_i \in \Theta_i \), \( \theta_i \sim_i \theta'_i \) implies \( f(\theta_i, \theta_{-i}) = f(\theta'_i, \theta_{-i}) \) for all \( \theta_{-i} \in \Theta_{-i} \).

Strong \( \text{dr}^\Gamma \)-measurability implies \( \text{dr}^\Gamma \)-measurability. If \( \Gamma \) is a static mechanism, then strong \( \text{dr}^\Gamma \)-measurability and \( \text{dr}^\Gamma \)-measurability are equivalent.

**Proposition 2** Let \( \Gamma^* \) be a mechanism such that \( \mathcal{H}^*_i(F^*_{+\infty}) = \mathcal{H}^*_i \) for all \( i \in I \) and suppose the environment satisfies the economic property. Then every ex-post incentive compatible social choice function that is strongly \( \text{dr}^\Gamma^* \)-measurable is robustly virtually implementable.

**Proof.** First, note some facts about the environment. a) of the following lemma follows directly from the finiteness of \( I, X \) and \( \Theta \). To see b) and c), choose \( \alpha \) sufficiently small such that b) holds for \( c_0 = 0 \) and then use finiteness of \( I \) and \( \Theta \).

**Lemma 1** There exist \( C_0 > 0 \) and, if the economic property holds, \( c_0 > 0 \) and \( \alpha \in (0, 1) \) such that

1. \( |u_i(y, \theta) - u_i(y', \theta)| \leq C_0 \) for all \( i \in I, \ y, y' \in Y \) and \( \theta \in \Theta \).
2. \( u_i(\alpha \bar{y} + (1 - \alpha)z_i + \sum_{j \neq i} z_j, \theta) > u_i(\bar{y} + \sum_{j \neq i} (\alpha \bar{y} + (1 - \alpha)z_j), \theta) + c_0 \) for all \( i \in I \) and \( \theta \in \Theta \).
3. \( u_i(z_i, \theta) > u_i(\alpha \bar{y} + (1 - \alpha)z_i, \theta) + c_0 \) for all \( i \in I \) and \( \theta \in \Theta \).

The next lemma, proved in appendix A, says that in any (dynamic) mechanism, the loss in utility from playing a strategy that is not \((k + 1)\)-rationalizable is uniformly bounded below, much like for static mechanisms.

**Lemma 2** For any mechanism \( \Gamma \) there exists \( \eta_{\Gamma} > 0 \) such that for any \( i \in I, \ k \in \mathbb{N}, \ (s_i, \theta_i) \notin F^k_i \) and \( \mu_i \in \Phi_i^k \) there are \( \mathcal{H} \in \mathcal{H}_i(s_i) \) and \( s'_i \in S_i(\mathcal{H}) \) such that

\[
U^\mu_i(s'_i, \theta_i, \mathcal{H}) > U^\mu_i(s_i, \theta_i, \mathcal{H}) + \eta_{\Gamma}.
\]

Next, given \( \Gamma^* = (H^*, (\mathcal{H}_i^*)_{i \in I}, P^*, C^*) \), let \( \delta > 0 \) and \( L \in \mathbb{N}\setminus\{0\} \) be such that

\[
\delta^2 C_0 < \delta \eta_{\Gamma^*} \quad (2)
\]

and

\[
\frac{1}{L} C_0 < \delta^2 \frac{1}{L} c_0, \quad (3)
\]

where \( C_0 \) and \( c_0 \) are the constants from lemma 1, and \( \eta_{\Gamma^*} \) the constant from lemma 2. We now define a mechanism \( \Gamma = (H, (\mathcal{H}_i)_{i \in I}, P, C) \) that will robustly \( \sqrt{2}(\delta + \delta^2) \)-implement \( f \): At first, each agent submits \( L \) times what his payoff type is. He can lie, and his \( l \)-th submission
can differ from is m-th submission. The agents do this simultaneously and their submission is not revealed to the other agents during the entire mechanism. Afterwards, the agents play \( \Gamma^* \). Formally, the set of histories is\(^\dagger\)

\[
H = \{ h \in \mathcal{F}; h \preceq (\theta_1, \ldots, \theta_I, h^*) \text{ for some } (\theta_i)_{i \in I} \in \Theta^L_I \times \cdots \times \Theta^L_I, h^* \in H^* \},
\]

where \( \mathcal{F} \) is the set of finite sequences with codomain \( A^* \cup \bigcup_{i \in I} \Theta^L_i \). The player function \( P : H \setminus T \rightarrow I \) is defined by \( P(\theta_1, \ldots, \theta_{i-1}) = i \) for \( i \in I \) and \( P(\theta_1, \ldots, \theta_I, h^*) = P^+(h^*) \) for \( (\theta_j)_{j \in I} \in \Theta^L_I \times \cdots \times \Theta^L_I \) and \( h^* \in H^* \). Agent \( i \)'s information sets are \( \mathcal{H}^0_i = \{ (\theta_1, \ldots, \theta_{i-1}); \theta_j \in \Theta^L_j \text{ for } j < i \} \) and

\[
[\theta_i, \mathcal{H}^*] = \{ (\theta_1, \ldots, \theta_i, h^*); \theta_j \in \Theta^L_j \text{ for } j \neq i, h^* \in \mathcal{H}^* \}, \quad \forall \theta_i \in \Theta^L_i, \mathcal{H}^* \in \mathcal{H}^*_i,
\]

so that \( \mathcal{H}_i = \{ \mathcal{H}^0_i \} \cup \{ [\theta_i, \mathcal{H}^*]; \theta_i \in \Theta^L_i, \mathcal{H}^* \in \mathcal{H}^*_i \} \). The agents' \( l \)-th submissions of their payoff types are used to determine almost \( 1 \)-th of the outcome via a direct mechanism, so that if all agents truthfully announce their payoff type \( L \) times the outcome of the mechanism will almost equal the outcome stipulated by \( f \). Since \( f \) is strongly \( \Gamma^* \)-measurable this is also true if the agents submit payoff types that are \( \Gamma^* \)-strategically indistinguishable from their true payoff types. The rest of the outcome provides the agents with incentives not to lie in their submissions of their payoff types (in the sense that they do not want to announce payoff types that are \( \Gamma^* \)-strategically distinguishable from their true payoff types), and is determined by the agents' play in \( \Gamma^* \), and to a smaller extent by some reward terms. More precisely, the outcome function \( C : T \rightarrow Y \) assigns the lottery

\[
C(h) = (1 - \delta - \delta^2) \frac{1}{L} \sum_{l=1}^{L} f(\theta^l) + \delta C^*(h^*) + \delta^2 \frac{1}{I} \sum_{i \in I} r_i(h),
\]

to terminal history \( h = ((\theta^1_i, \ldots, \theta^L_i)_{i \in I}, h^*) \), where agent \( i \)'s reward \( r_i(h) \) is defined by

\[
r_i((\theta^1_i, \ldots, \theta^L_i)_{i \in I}, h^*) = \begin{cases} 
\bar{y} & \text{if } h^* \notin H^*(R_{i,\infty}^*(\theta^1_i)) \\
\alpha \bar{y} + (1 - \alpha) z_i & \text{if } h^* \in H^*(R_{i,\infty}^*(\theta^1_i)) \text{ and } \exists m \in \{2, \ldots, L\} : \theta^m_i \neq \theta^1_i \text{ and } \forall l \in \{2, \ldots, m-1\} : \theta^l_{-i} = \theta^1_{-i} \\
z_i & \text{otherwise}
\end{cases}
\]

where \( \alpha \) is the constant from lemma 1. The next lemma shows that if \( h = ((\theta^1_i, \ldots, \theta^L_i)_{i \in I}, h^*) \) is admitted in \( \Gamma \) by a strategy profile that is strongly rationalizable for \( \theta \), then \( h^* \) is admitted in \( \Gamma^* \) by a strategy profile that is strongly rationalizable for \( \theta \): \( \Gamma \) strategically distinguishes any payoff type profiles that \( \Gamma^* \) strategically distinguishes. The term \( r_i(h) \) punishes \( i \) with \( \bar{y} \)

\(^\dagger\)For a finite sequence \( h \) with codomain \( A \) and \( a_1, \ldots, a_n \in A \), let \( (a_1, \ldots, a_n, h) \) denote the finite sequence \( g \) with codomain \( A \cup A' \) and length \( n + l_h \) such that \( g(k) = a_k, k = 1, \ldots, n, \) and \( g(n+k) = h(k), k = 1, \ldots, l_h \).
Lemma 3

For each \( \phi_i \) of his payoff type, and is detected. If he does not lie or his lie goes undetected, he is rewarded with \( \alpha y + (1 - \alpha)z_i \) or with \( z_i \). He gets the less preferred reward \( \alpha y + (1 - \alpha)z_i \) if he is one of the agents to deviate “first” from his first announcement \( \theta_1^i \) in one of his later announcements \( \theta_i \), and the more preferred reward \( z_i \) otherwise.

Let \( \varphi_i : S_i \rightarrow S_i^* \) be such that

\[
\varphi_i(s_i)(\mathcal{H}^*) = s_i([s_i(\mathcal{H}_i^0), \mathcal{H}^*]), \quad \forall \mathcal{H}^* \in \mathcal{H}_i^*.
\]

\( \varphi_i(s_i) \) is the strategy induced by the “\( \Gamma^* \)-part” of \( s_i \) that is not prevented by \( s_i \) itself. Moreover, let \( \phi^0_i : S_i^* \rightarrow S_i \) be defined by \( \phi^0_i(s_i^*)(\mathcal{H}_i^0) = (\theta_i, \ldots, \theta_i) \) and

\[
\phi^0_i(s_i^*)(\theta_i, \mathcal{H}^*) = s_i^*(\mathcal{H}^*), \quad \forall \theta_i \in \Theta_i, \mathcal{H}^* \in \mathcal{H}_i^*.
\]

**Lemma 3**

For each \( k \in \mathbb{N}, i \in I, \theta_i \in \Theta_i, s_i^* \in S_i^* \) and \( s_i \in S_i \),

a) \( s_i^* \in R_i^{k,\infty}(\theta_i) \) implies \( \phi^0_i(s_i^*) \in R_i^k(\theta_i) \) and

b) \( s_i \in R_i^k(\theta_i) \) implies \( \varphi_i(s_i) \in R_i^{*,k}(\theta_i) \).

We prove lemma 3 by induction on \( k \), and are actually only interested in part b). To get an intuition for why b) holds, note that at the time an agent reaches his decision nodes in the \( \Gamma^* \)-part of \( \Gamma \) he has already committed to the \( L \) announcements of his type and can only influence two components of \( \Gamma \)’s outcome: \( \delta C^s(h^*) \) and \( \delta \sum_{i \in I} r_i(h) \). Take a \( s_i \in R_i^k(\theta_i) \), \((s_i, \theta_i) \in \rho_i(\mu_i)\), and suppose b) holds for all \( k' < k \). Project \( \theta_i \)’s belief \( \mu_i \) to a belief in \( \Gamma^* \). If \( \varphi_i(s_i) \) is not a sequential best response to the projected belief for \( \theta_i \) in \( \Gamma^* \), then by lemma 2 there must be a strategy in \( S_i^* \) that promises at least \( \eta_{\mathcal{H}^*} \) more expected utility at some decision node. But then, playing that superior strategy in the \( \Gamma^* \)-part of \( \Gamma \) leads to an expected utility gain that by (2) cannot be offset even by the largest conceivable loss the superior strategy can cause in the reward terms. This contradicts \( s_i \in R_i^k(\theta_i) \).

This argument is not yet complete, however: there could be a decision node \( \mathcal{H}_i \) that in \( \Gamma^* \) is admitted by \( (k - 1) \)-rationalizable strategies of \( i \)’s opponents, and in \( \Gamma \) by say \( (k - 2) \)-rationalizable but not by \( (k - 1) \)-rationalizable strategies of \( i \)’s opponents. In that case, projecting \( \mu_i \in \Phi_{i}^{k-1} \) to \( \Delta \mathcal{H}_i^*(\Sigma_{\mathcal{H}_i}) \) does not necessarily yield an element of \( \Phi_{i}^{*,k-1} \): At \( \mathcal{H}_i \), the projected belief puts probability one on \( F_i^{*,k-2} \), but not necessarily probability one on \( F_i^{*,k-1} \). That is where a) comes into play: Because \( \mathcal{H}_i^*(\Sigma_{\mathcal{H}_i}) = \mathcal{H}_i^* \) for all \( i \in I \) all decision nodes in \( \Gamma^* \) are reached by some strongly rationalizable strategy profile of some payoff type profile. Thus if a) is true for all \( k' < k \), all decision nodes in \( \Gamma \) are reached by some \( (k - 1) \)-rationalizable strategy profile of some payoff type profile, and the situation just described cannot arise. We relegate a formal proof of lemma 3 to appendix A.

\(^1\)\( \varphi_i \) is surjective but not necessarily injective, while \( \phi^0_i \) is injective but not necessarily surjective.
The next lemma shows that as a consequence of lemma 3 b), agents will never announce payoff types that are $\Gamma^*$-strategically distinguishable from their true payoff types — in that sense, they will never lie in their announcements of their payoff types.

**Lemma 4** $(s, \theta) \in F_{\infty}^\infty$ implies $s_i(H_i^0)^l \sim_{\Gamma^*} \theta_i$ for all $l \in \{1, \ldots, L\}$ and all $i \in I$.

**Proof.** Let $i \in I$, $\mu_i \in \Phi_i^\infty$, and $(s_i, \theta_i) \in \rho_i(\mu_i)$. Suppose $s_i(H_i^0)^1 \sim_{\Gamma^*} \theta_i$. We will argue that the strategy $s_i' \in S_i$ that equals $s_i$ except that $s_i'(H_i^0)^1 = \theta_i$ promises payoff type $\theta_i$ a strictly higher expected utility at $H_i^0$ under $\mu_i$ than $s_i$. First, agent $i$ is certain that his lie $s_i(H_i^0)^1$ will get detected since it is $\Gamma^*$-strategically distinguishable from his true type: Suppose not, then he must put strictly positive probability on some $(s_i, \theta_i, \theta_i) \in F_{\infty}^\infty$ which satisfies $h^* \equiv \zeta^*(\varphi_1(s_1), \ldots, \varphi_L(s_L)) \in H^*(R^*_{\infty}(s_i(H_i^0)^1)).$ But then there must exist $\hat{s}_i^* \in R^*_{\infty}(s_i(H_i^0)^1)$ such that $\zeta^*(\hat{s}_i^*, (\varphi_j(s_j))_{j \neq i}) = h^*$. By lemma 3 b), $(\varphi_j(s_j), \theta_j) \in F_{\infty}^\infty$ for $j \in I$, and hence $h^* \in H^*(R^*_{\infty}(\theta_j))$ for all $j \in I$. That implies $(\theta_i, \theta_i) \sim_{\Gamma^*} (s_i(H_i^0)^1, \theta_i)$ and hence $s_i(H_i^0)^1 \sim_{\Gamma^*} \theta_i$. Contradiction. Therefore, $s_i$ leads to $r_i(\zeta(s_i, s_{-i})) = \bar{y}$ for any $s_{-i}$ that agent $i$ expects with strictly positive probability, while, again by lemma 3 b), $r_i(\zeta(s_i', s_{-i})) \in \{\alpha \bar{y} + (1 - \alpha)z_i, z_i\}$ for all $s_{-i} \in S_{-i}$. Second, if agent $i$ strictly prefers $r_j(\zeta(s_i, s_{-i}))$ over $r_j(\zeta(s_i', s_{-i}))$ for some $s_{-i} \in S_{-i}$, then $r_j(\zeta(s_i, s_{-i})) = \alpha \bar{y} + (1 - \alpha)z_j$ and $r_j(\zeta(s_i', s_{-i})) = z_j$. This is because $s_i$ and $s_i'$ agree in their $\Gamma^*$-part and thus $r_j(\zeta(s_i, s_{-i})) = \bar{y}$ if and only if $r_j(\zeta(s_i', s_{-i})) = \bar{y}$. Therefore, by playing $s_i'$ instead of $s_i$, agent $i$ incurs a loss (if at all) of less than $\frac{1}{\Lambda}C_0$ in $L(1 - \delta - \delta^2)\frac{1}{\Lambda}f((\hat{s}_i(H_i^0)^1)_{i \in I})$ but by lemma 1 b) gains at least $\delta^2 \frac{1}{2}C_0$ in $\frac{\delta^2}{2} \sum_{i \in I} r_i(\zeta(\hat{s}_i)).$ By (3), the gain outweighs the potential loss. Hence $s_i(H_i^0)^1 \sim_{\Gamma^*} \theta_i$.

It now suffices to show that $s_i(H_i^0)^l = s_i(H_i^0)^1$ for all $l \in \{2, \ldots, L\}$, $(s_i, \theta_i) \in F_{\infty}^\infty$ and $i \in I$. Suppose this is false. Pick $i \in I$, $(s_i, \theta_i) \in \rho_i(\mu_i)$ and $\mu_i \in \Phi_i^\infty$ such that $s_i(H_i^0)^m \neq s_i(H_i^0)^1$, where $m$ is the minimal element in $\{2, \ldots, L\}$ for which there are $i \in I$ and $(s_i, \theta_i) \in F_{\infty}^\infty$ such that $s_i(H_i^0)^m \neq s_i(H_i^0)^1$.

As a first case, suppose that $\mu_i((s_{-i}, \theta_{-i})|H_i^0) > 0$ implies $s_j(H_j^0)^l = s_j(H_j^0)^1$ for all $l \in \{2, \ldots, L\}$ and all $j \neq i$. Then, strategy $s_i'$ defined by $s_i'(H_i^0)^l = (\theta_i, \ldots, \theta_i)$ and $s_i'(\{\theta_i, \theta_i^*\}) = s_i'(s_i(H_i^0)^1)$ for all $\theta_i, \theta_i^*$ gives strictly higher expected utility at $H_i^0$ than $s_i$ for a payoff type $\theta_i$ with beliefs $\mu_i$, contradicting $(s_i, \theta_i) \in \rho_i(\mu_i)$: Because $f$ is epIC and strongly $\Gamma^*$-measurable, $s_i'$ maximizes the expected utility from $(1 - \delta - \delta^2)\frac{1}{2} \sum_{i=1}^L f((\hat{s}_i(H_i^0)^1)_{i \in I})$ in $S_i$. Strategies $s_i$ and $s_i'$ yield the same expected utility from $\delta C^*(\zeta^*(\varphi_1(s_1), \ldots, \varphi_L(s_L)))$. By lemma 3 b) $r_i(\zeta(s_i', s_{-i})) = z_i$, while $r_i(\zeta(s_i)) \in \{\bar{y}, \alpha \bar{y} + (1 - \alpha)z_i\}$ for any strategy profile $s_{-i}$ that $i$ expects to be played with strictly positive probability. Since $s_i$ prescribes that agent $i$'s $m$-th announcement deviates from his first announcement, while $s_i'$ prescribes no deviation at all, $s_i'$ must lead to a weakly better outcome from $r_j$ than $s_i$: for any $s_{-i}$ that $i$ expects to be played with strictly positive probability and any $j \neq i$, $r_j(\zeta(s_i', s_{-i})) = \bar{y}$ if and only if $r_j(\zeta(s_i, s_{-i})) = \bar{y}$ because $s_i$ and $s_i'$ agree in their $\Gamma^*$-part, and if $r_j(\zeta(s_i', s_{-i})) = z_j$ then
By lemma 1 c) and (3), the expected utility difference between rather than

\[ r_j(\zeta(s_i, s_{-i})) = z_j. \]

As a second case, suppose that

\[ n = \min \left\{ l \in \{2, \ldots, L\}; \exists(s_{-i}, \theta_{-i}) \in \Sigma_{-i}, j \neq i : \mu_i((s_{-i}, \theta_{-i})|\mathcal{H}^\theta_{j}) > 0 \land s_j(\mathcal{H}^\theta_{j}) \neq s_j(\mathcal{H}^\theta_{i}) \right\} \]

is well-defined. \( n \geq m \) is the smallest \( l \) for which \( i \) expects some \( j \neq i \) to deviate from his first announcement. Define strategy \( s'_i \) by \( s'_i(\mathcal{H}^\theta_{j}) = (\theta_{j}, \ldots, \theta_{j}, s_i(\mathcal{H}^\theta_{i})^{n+1}, \ldots, s_i(\mathcal{H}^\theta_{i})^{L}) \) and \( s'_i([\theta_{j}, \mathcal{H}^\theta_{j}]) = s_i([s_i(\mathcal{H}^\theta_{i}), \mathcal{H}^\theta_{i}]) \) for all \( [\theta_{j}, \mathcal{H}^\theta_{j}] \). Then \( s'_i \) gives strictly higher expected utility at \( \mathcal{H}^\theta_{j} \) than \( s_i \) for \( \theta_i \) under \( \mu_i \), contradicting \( (s_i, \theta_i) \in \rho_i(\mu_i) \): Because \( f \) is epiIC and strongly \( \text{dr}^{\Gamma^*} \)-measurable, \( s'_i \) maximizes the expected utility from \( (1 - \delta - \delta^2) \frac{1}{2} \sum_{l=1}^{n-1} f((\hat{s}_i(\mathcal{H}^\theta_{j}))_{l \in I}) \) in \( S_i \). \( s'_i \) yields the same expected utility as \( s_i \) from \( (1 - \delta - \delta^2) \frac{1}{2} \sum_{l=n+1}^{L} f((\hat{s}_i(\mathcal{H}^\theta_{j}))_{l \in I}) + \delta C^*(\zeta^*(\varphi_1(\hat{s}_1), \ldots, \varphi_L(\hat{s}_L))). \) Agent \( i \) expects that with probability \( p \in (0, 1] \), there is \( j \neq i \) who submits \( s_j(\mathcal{H}^\theta_{j}) \neq s_j(\mathcal{H}^\theta_{i}) \). Thus for strategy-payoff type profiles \( (s_{-i}, \theta_{-i}) \) that total probability mass \( p \), agent \( i \) expects \( r_i(\zeta(s'_i, s_{-i})) = z_i \) and \( r_i(\zeta(s)) \in \{ \alpha \bar{y} + (1 - \alpha)z_i, \bar{y} \} \).

By lemma 1 c) and (3), the expected utility difference between \( z_i \) and \( \alpha \bar{y} + (1 - \alpha)z_i \) (and therefore, the expected utility difference between \( z_i \) and \( \bar{y} \), as well) strictly outweighs the possible expected utility loss from the term \( (1 - \delta - \delta^2) \frac{1}{2} f((\hat{s}_i(\mathcal{H}^\theta_{j}))_{l \in I}) \). Since by playing \( s'_i \) agent \( i \) deviates “later” from his first announcement of his payoff type than by playing \( s_i \), agent \( i \) weakly prefers \( r_j(\zeta(s'_i, s_{-i})) \) over \( r_j(\zeta(s_i, s_{-i})) \) for any \( j \neq i \) and any \( s_{-i} \in S^\theta_{-i} \).

With probability \( 1 - p \), \( i \) expects no \( j \) to deviate in his \( n \)-th announcement of his payoff type, in which case \( s'_i \) is expected to lead to no worse reward \( r_i \) than \( s_i \) (\( s_i \) deviates in \( m \)-th submission and hence leads at best to \( \alpha \bar{y} + (1 - \alpha)z_i \); \( s'_i \) leads to no worse than \( \alpha \bar{y} + (1 - \alpha)z_i \)). As for the rewards \( r_j \), \( j \neq i \), \( s_i \) and \( s'_i \) lead to \( \bar{y} \) in exactly the same cases, and playing \( s'_i \) rather than \( s_i \) cannot decrease the probability with which \( i \) expects \( r_j \) to equal \( \alpha \bar{y} + (1 - \alpha)z_j \) versus \( z_j \).

By strong \( \text{dr}^{\Gamma^*} \)-measurability of \( f \), \( \theta_j \sim^{\Gamma^*} \theta_j \) for all \( j \in I \) implies \( f(\theta) = f(\theta') \), for any \( \theta, \theta' \in \Theta \). Therefore, by lemma 4, \( (s, \theta) \in F^\infty \) implies

\[ ||C(\zeta(s)) - f(\theta)|| \leq \delta ||C^*(\zeta^*(\varphi(s))) - f(\theta)|| + \delta^2 \frac{1}{L} \sum_{i \in I} r_i(\zeta(s)) - f(\theta)|| \leq \sqrt{2}(\delta + \delta^2), \]

and \( \Gamma \) robustly \( \sqrt{2}(\delta + \delta^2) \)-implements \( f \). Since \( \delta \) can be chosen arbitrarily small, \( f \) is rv-implmentable. This completes the proof of proposition 2.

Proposition 2 is not an exact converse of proposition 1. First, our mechanism uses small punishments and rewards whose existence are guaranteed by the economic property. This is much like in Abreu and Matsushima (1992a, b), Artemov, Kunimoto, and Serrano (2009) and BM. Second, some new difficulties arise in our case because \( \Gamma^* \) can be an dynamic mechanism (all of the just mentioned papers only use static mechanisms).
Strong dr-measurability and dr-measurability are not equivalent in our case. While example 4.1 shows that strong dr-measurability is not necessary for a social choice function to be rv-implementable in general, strong dr-measurability makes sure that “epIC is insensitive to strategically indistinguishable lies of others”. That is, strong dr-measurability makes sure telling the truth is a best response for $i$ in the direct mechanism associated with $f$ if $i$ expects others to lie strategically indistinguishable.

Moreover, we do not know if there exists a “maximally revealing mechanism” that strategically distinguishes all payoff type profiles that are strategically distinguishable by some mechanism. Hence, we assume strong dr-measurability with respect to some mechanism.

The following example demonstrates that strong dr-measurability is in general not necessary for rv-implementability.

**Example 4.1** There are two agents $i \in \{1, 2\}$ with two payoff types each, $\Theta_i = \{\theta_i, \theta_i'\}$, and three pure outcomes, $X = \{x, y, z\}$. Player 1 prefers “not $z$” when he is of payoff type $\theta_1$ and $z$ when he is of payoff type $\theta_1'$:

$$u_1(x, \theta_1, \cdot) = u_1(y, \theta_1, \cdot) > u_1(z, \theta_1, \cdot)$$
$$u_1(z, \theta_1', \cdot) > u_1(x, \theta_1', \cdot) = u_1(y, \theta_1', \cdot)$$

Player 2 is indifferent between all outcomes unless the payoff type profile is $(\theta_1, \theta_2)$, in which case he favors $x$, or $(\theta_1, \theta_2')$, in which case he favors $y$:

$$u_2(x, \theta_1, \theta_2) > u_2(y, \theta_1, \theta_2) = u_2(z, \theta_1, \theta_2)$$
$$u_2(y, \theta_1, \theta_2') > u_2(x, \theta_1, \theta_2') = u_2(z, \theta_1, \theta_2')$$
$$u_2(x, \theta_1', \cdot) = u_2(y, \theta_1', \cdot) = u_2(z, \theta_1', \cdot)$$

Clearly $(\theta_1', \theta_2) \sim (\theta_1', \theta_2')$,\(^\dagger\) and therefore $\theta_2 \sim_2 \theta_2'$. The social choice function $f : \Theta \to \Delta(X)$ given in figure 3 is not strongly dr-measurable, but rv-implementable via mechanism $\Gamma$.

In section 5, we will consider private consumption environments that satisfy the economic property. We will present a mechanism $\Gamma$ that in private consumption environments with generic valuation functions strategically distinguishes all payoff type profiles, and in which every history can be rationalized reached by some payoff type profile. In this case, dr- and strong dr$^\Gamma$-measurability coincide (both are trivially satisfied by every social choice function),

\[^\dagger\text{Take any mechanism } \Gamma \text{ and pick an arbitrary } s'_1 \in R^n_1(\theta'_1). \text{ Let } \delta_{(s'_1, \theta'_1)} \in \Delta(\Sigma_1) \text{ denote the point belief in } (s'_1, \theta'_1) \text{ and let } \mu'_2 \in \Phi^n_2. \text{ Define } \mu'_2 : 2^{\Sigma_1} \times \mathcal{H}_2 \to [0, 1] \text{ by } \mu'_2(\cdot|\mathcal{H}) = \delta_{(s'_1, \theta'_1)} \text{ for } \mathcal{H} \in \mathcal{H}_2(s'_1) \text{ and } \mu'_2(\cdot|\mathcal{H}) = \mu'_2(\cdot|\mathcal{H}) \text{ for } \mathcal{H} \notin \mathcal{H}_2(s'_1). \text{ Note that } \mu'_2 \text{ is a CPS, and, since } \mu'_2 \in \Phi^n_2 \text{ and all the mass of } \delta_{(s'_1, \theta'_1)} \text{ concentrates on a strongly rationalizable strategy-payoff type pair, } \mu_2 \in \Phi^n_2. \text{ Let } s'_2 \in r_2(\theta_2, \mu_2), \text{ then there exists } s_2 \in r_2(\theta_2, \mu_2) \text{ such that } s_2|\mathcal{H}_2(s'_2) = s'_2|\mathcal{H}_2(s'_1). \text{ In summary, } (s'_1, \theta'_1) \in F^n_1, (s_2, \theta_2), (s'_2, \theta'_2) \in F^n_2 \text{ and } \zeta(s'_1, s_2) = \zeta(s'_1, s'_2).\]
the additional assumptions of proposition 2 have no bite and an exact characterization of rv-implementation obtains: a social choice function is robustly virtually implementable if and only if it is ex-post incentive compatible (corollary 1).

5 Strategic Distinguishability

We focus on an economic environment in which there is a single good that is to be allocated to one of the agents. The set of outcomes is

$$Y = \{(q, t) = (q_1, \ldots, q_I, t_1, \ldots, t_I) \in [0, 1]^I \times [-B, B]^I \mid \sum_{i \in I} q_i \leq 1\},$$

where $B > 0$, $q_i$ is interpreted as the probability that agent $i$ gets the good, and $t_i$ as a monetary transfer to agent $i$. Agent $i$'s utility is given by $u_i((q, t), \theta) = v_i(\theta)q_i + t_i$, where $v_i : \Theta \to \mathbb{R}$ is his valuation function. We call such an environment a private consumption environment.

Remark 1 The outcome space $Y$ is the “reduced form” of a space of lotteries over a finite set of pure outcomes in the following sense. Let the set of pure outcomes be

$$\bar{X} = \{(\bar{q}, \bar{t}) = (\bar{q}_1, \ldots, \bar{q}_I, \bar{t}_1, \ldots, \bar{t}_I) \in \{0, 1\}^I \times \{-B, B\}^I \mid \sum_{i \in I} \bar{q}_i \leq 1\},$$

and let agent $i$’s utility function be

$$\bar{u}_i : \bar{X} \times \Theta \to \mathbb{R}, ((\bar{q}, \bar{t}), \theta) \mapsto v_i(\theta)\bar{q}_i + \bar{t}_i$$

where $v_i : \Theta \to \mathbb{R}$ is the valuation function introduced above. Preferences over lotteries in the outcome space $\bar{Y} = \Delta(\bar{X})$ are expected utility preferences with respect to $\bar{u}_i$. Then there is a utility-preserving surjection $g$ from $\bar{Y}$ to $Y$ (for details, see appendix B). Therefore, the results of section 4 apply to the current section.

Note that private consumption environments satisfy the economic property.

\footnote{The results of this section immediately generalize to the case of multiple goods (including the case of complements and substitutes). Our sufficiency proof, however, exploits the privacy of consumption, i.e. the fact that $i$ is indifferent between $(q, q_{-i})$ and $(q, q_{-i}')$ (given some transfers $t$).}
5.1 Sufficient Conditions for Strategic Distinguishability

Proposition 3 is the main result of this section: Under a weak sufficient condition, all payoff type profiles can be strategically distinguished in private consumption environments. Given a valuation function \( v \), let \( V_i(\theta_i) = \{ v_i(\theta_i, \theta_{-i}) | \theta_{-i} \in \Theta_{-i} \} \) be the set of \( \theta_i \)'s expected valuations that can arise from point beliefs, or the set of \( \theta_i \)'s ex-post valuations.

**Proposition 3** If a private consumption environment satisfies \( V_i(\theta_i) \cap V_i(\theta'_i) = \emptyset \) for all \( i \in I, \theta_i, \theta'_i \in \Theta_i, \theta_i \neq \theta'_i \), then there is a mechanism \( \Gamma \) that strategically distinguishes all payoff type profiles and in which every non-terminal history is admitted by strongly rationalizable strategies of some payoff type profile. That is, there is a mechanism \( \Gamma \) such that both \( \theta \nleq^\Gamma \theta' \) for all \( \theta, \theta' \in \Theta, \theta \neq \theta' \), and \( H \setminus T = H(F^\infty) \).

**Proof.** For each agent \( i \in I \) we define some “options” \( o_i^m = (q_i^m, t_i^m) \in [0, 1] \times [-B, B] \), consisting of a probability \( q_i^m \) and a transfer \( t_i^m \). These options will later be used in defining the outcome function of \( \Gamma \). As figure 4 indicates, a line in a diagram with agent \( i \)'s utility \( u_i = v_i q_i + t_i \) on the vertical and agent \( i \)'s valuation \( v_i \) on the horizontal axis uniquely determines an option.

![Figure 4: An arbitrary option for agent i](image)

For \( i \in I \), let \( V_i = \bigcup_{\theta_i \in \Theta_i} V_i(\theta_i) \) be the set of agent \( i \)'s ex-post valuations, let \( v_i^n \) denote the \( m \)-th smallest element of \( V_i \) and let \( n_i = \# V_i \) (so that \( V_i = \{ v_i^1, \ldots, v_i^n \} \) and \( v_i^1 < \ldots < v_i^n \)). Let \( \pi_i \) be a selection of the correspondence that assigns to each \( v_i \in V_i \) the set of payoff type profiles \( \{ \theta \in \Theta; v_i(\theta) = v_i \} \), and let \( v_i^0 = v_i^1 - 1 \). Define an option for each element of \( V_i \) as shown in figure 5 (the utility scale in figure 5 is such that \( o_i^m \in [0, \frac{1}{2}] \times [-B, B] \) for all \( m \in \{ 1, \ldots, n_i \} \), plus the option \( o_i^0 = (0, 0) \). The following lemma is obvious from figure 5.
A formal proof and a formal definition of the options can be found in appendix A.

Lemma 5 For $i \in I$ and $m \in \{1, \ldots, n_i\}$,

a) $u_i(q^m_i, \pi_i(v^l_i)) < u_i(q^0_i, \pi_i(v^l_i)) = 0$ for $l \in \{1, \ldots, m - 1\}$,

b) $u_i(q^m_i, \pi_i(v^l_i)) > u_i(q_i^l, \pi_i(v^m_i))$ for all $l \in \{0, \ldots, n_i\} \setminus \{m\}$.

We proceed to define the mechanism $\Gamma = (H, (K_i)_{i \in I}, P, C)$.$^1$ With each $m_i \in \{1, \ldots, n_i\}$ we associate the unique $\tau_i(m_i) \in \Theta_i$ for which there exists $\theta_{-i} \in \Theta_{-i}$ such that $v_i(\tau_i(m_i), \theta_{-i}) = v^m_i$, and, for all $m \in \prod_{i \in I} \{1, \ldots, n_i\}$ and all $i \in I$, let $m_i = (m_1, \ldots, m_i)$ and

$$V_i(\theta_i||m||i-1) = \{v_i(\theta_i, \theta_{-i})|\theta_j = \tau_j(m_j)\text{ for } j < i, \theta_j \in \Theta_j\text{ for } j > i\}$$

be the set of $\theta_i$'s valuations that can arise from point beliefs when $i$ knows the payoff types of the agents $j < i$ to be $\tau_1(m_1), \ldots, \tau_{i-1}(m_{i-1})$. The set of histories is

$$H = \{h \in \mathcal{H}; \exists m \in \prod_{i \in I} \{1, \ldots, n_i\} : h \preceq m \text{ and } \forall i \in I \exists \theta_i \in \Theta_i : v^m_i \in V_i(\theta_i||m||i-1)\}$$

where $\mathcal{H}$ is the set of finite sequences with codomain $\{1, \ldots, \max_{i \in I} n_i\}$. For $i \in I$, let $>^\mathbb{N}$ order lexicographically: $m_i > m_i'$ if there is $j \in \{1, \ldots, i\}$ such that $m_k = m'_k$ for $k < j$ and $m_j > m'_j$. For $m \in H = I$, $i \in I$ and $j \in \{0, \ldots, I\}$, let $\bar{m}(m_i)$ denote the largest element $m'$

\(^1\text{We assume that } \#\Theta_i > 2 \text{ for all } i \in I. \text{ In this case, the game form we define cannot have trivial decision nodes and is thus a mechanism (see definition 2). The case in which } \Theta_i \text{ is a singleton for some } i \in I \text{ can be easily accommodated at the cost of additional notation.}\)

\(^2\text{m}_0 = \emptyset \text{ and } V_i(\theta_1||m_0) = V_i(\theta_1)\).

Figure 5: The options $q^m_i$ for agent $i$
in $H^I$ such that $m'_i = m_i$, let $\tilde{m}(m|_0) = \max H^I$, and $\tilde{m}_i(m|_j) = [\tilde{m}(m|_j)]_i$. Let $\underline{m}(m|_i)$ denote the smallest element $m'$ in $H^I$ such that $m'|_i = m_i$, let $\underline{m}(m|_0) = \min H^I$, and $\underline{m}_i(m|_j) = [\underline{m}(m|_j)]_i$. If $m_i \neq \underline{m}_i(m|_{i-1})$, let $\underline{m}_i(m|_j)$ denote the largest element $m'$ in $H^I$ such that $m'|_{i-1} = m_{i-1}$ and $m'_i < m_i$, let $\underline{m}_i(m|_j) = [\underline{m}(m|_j)]_i$.

The players move sequentially, that is, the player function is $P : H^I \rightarrow I$ such that $P(m_1, \ldots, m_{i-1}) = i$ for all $i \in I$ and all $(m_1, \ldots, m_{i-1}) \in H^I$. Every action taken is immediately visible to all other agents, $\mathcal{H}_i = \{m \in H_i \}$. Finally, the outcome function is $C : T \rightarrow Y$ such that $C(m) = (o_1, \ldots, o_I)$ where

$$o_i = \begin{cases} o^\hat{m}_i & \text{if } (\tau_1(m_1), \tau_i(m_i)) = \hat{v}_i \tilde{m}_i, \tilde{m}_i \in \{m_i, \ldots, n_i\} \text{ and } m_i \neq \underline{m}_i(m|_{i-1}) \\ o_{i+1} & \text{if } m_i = \underline{m}_i(m|_{i-1}) \\ o_i & \text{otherwise} \end{cases}$$

Let $\mathcal{F}_i = \{(s_i, \theta_i) \in \Sigma_i \forall m \in H^I : v_i^s(m|_{i-1}) \in V_i(\theta_i||m|_{i-1})\}$. Lemma 6 shows that $\mathcal{F} \subseteq F^\infty$, which implies that $H^I = H(F^\infty)$.

**Lemma 6** \(\mathcal{F} \subseteq F^\infty\).

*Proof.* Obviously, $\mathcal{F} \subseteq F^0$. Now suppose $\mathcal{F} \subseteq F^k$, \(k \in \mathbb{N}\), and let $i \in I$, \((s_i, \theta_i) \in \mathcal{F}_i\). For each $m|_{i-1} \in H_i$ let $(\theta_{i+1}^{|m|_{i-1}}, \ldots, \theta_{j}^{|m|_{i-1}}) \in \Theta_{i+1} \times \ldots \times \Theta_j$ be such that $v_i^s_j(m|_{j-1}) = v_i(\tau_1(m_1), \ldots, \tau_j(m_j))$, $\theta_i, \theta_{i+1}^{|m|_{i-1}}, \ldots, \theta_{j}^{|m|_{i-1}}$, and choose $s_{j}^{|m|_{i-1}} \in S_j$, $j \neq i$, such that both $(s_j^{|m|_{i-1}}, \tau_j(m_j)) \in \mathcal{F}_j$ and $s_{j}^{|m|_{i-1}} \{m|_{j-1}\} = m_j$ if $j < i$ and $(s_j^{|m|_{i-1}}, \theta_j^{|m|_{i-1}}) \in \mathcal{F}_j$ if $j > i$. Let $\mu_i : 2^{\Sigma_i \times \mathcal{H}_i} \rightarrow [0, 1]$ be such that

$$\mu_i((s_j^{|m|_{i-1}}, \tau_j(m_j)_{j<i}, s_j^{|m|_{i-1}}, \theta_j^{|m|_{i-1}}_{j>i}) \{m|_{i-1}\}) = 1 \quad \forall \{m|_{i-1}\} \in \mathcal{H}_i$$

and $\mu_i(\emptyset) = \mu_i(\{\tilde{m}|_{i-1}\})$ for an arbitrarily chosen $\{\tilde{m}|_{i-1}\} \in \mathcal{H}_i$. $\mu_i$ is a CPS, and by the induction hypothesis, $\mu_i \in \Phi^k$. Now take any $\{m|_{i-1}\} \in \mathcal{H}_i(s_i)$, then $U_i^\mu(s_i, \theta_i, \{m|_{i-1}\}) = v_i^s_j(m|_{j-1}) \tilde{d}_i^{s} + t_i^s(m|_{i-1})$. For any $s_i^j \in S_i(\{m|_{i-1}\})$ there is a $l \in \{0, \ldots, n_i\}$ such that $U_i^\mu(s_i^j, \theta_i, \{m|_{i-1}\}) = v_i^s_j(m|_{j-1}) \tilde{d}_i^{s} + t_i^s$. Hence by b) of lemma 5, $U_i^\mu(s_i^j, \theta_i, \{m|_{i-1}\}) \geq U_i^\mu(s_i^j, \theta_i, \{m|_{i-1}\})$. Therefore $(s_i, \theta_i) \in \rho_i(\mu_i) \subseteq F_i^{k+1}$.

To conclude the proof, we show that $F^\infty \subseteq \mathcal{F}$. For $m \in N^I$ let

$$\mathcal{F}(m) = \{(s_i, \theta) \in \Sigma; \forall i \in I \forall m \in H^I : (s_i \{m|_{i-1}\} = m_i \text{ and } m_i \geq m|_i) \implies \theta_i = \tau_i(m_i)\}.$$  

**Lemma 7** We have

a) $F^\infty \subseteq \mathcal{F}(\tilde{m}(n_1))$.

b) If $F^\infty \subseteq \mathcal{F}(m_1, \ldots, m_I)$, $m \in H^I$, then $F^\infty \subseteq \mathcal{F}(m_1, \ldots, m_{I-1}, 1)$. 

26
c) If \( F^\infty \subseteq \mathcal{F}(m_1, \ldots, 1), m_i \in H^{=i} \) and \( m_i \neq m_i^+(m_{i-1}) \), then \( F^\infty \subseteq \mathcal{F}(m_i^+(m_{i-1})) \).

Proof. a) Suppose \( (s_1, \theta_1) \in \rho_1(\mu_1), \mu_1 \in \Phi_1^\infty \), is such that \( s_1\{\emptyset\} = n_1 \). Let \( s'_1 \in S_1 \) be such that \( s'_1\{\emptyset\} = 1 \). There is \( \lambda \in \Delta(\{0, n_1\} \times \Theta_{-1}) \) such that 
\[
U^\mu_1(s_1, \theta_1, \emptyset) = \sum_{l \in \{0, n_1\}, \theta_{-1} \in \Theta_{-1}} (v(\theta) q^1_l + t^1_l) \lambda\{(l, \theta_{-1})\}.
\]
Since \( U^\mu_1(s'_1, \theta_1, \emptyset) \geq v^1 q^1_l + t^1_l > 0 \), \((s_1, \theta_1) \in \rho_1(\mu_1)\) implies that 
\[
\sum_{l \in \{0, n_1\}, \theta_{-1} \in \Theta_{-1}} (v(\theta) q^1_l + t^1_l) \lambda\{(l, \theta_{-1})\} > 0.
\]
By lemma 5 a), this is possible only if \( \tau_1(n_1) = \theta_1 \).

Suppose now that for \( j < i \), \((s_j, \theta_j) \in F_j^\infty \) and \( s_j\{\tilde{m}(n_1)|_{j-1}\} = \tilde{m}_j(n_1) \) imply \( \tau_j(\tilde{m}_j(n_1)) = \theta_j \). Suppose further that \((s_i, \theta_i) \in \rho_i(\mu_i), \mu_i \in \Phi_i^\infty \), is such that \( s_i\{\tilde{m}(n_1)|_{i-1}\} = \tilde{m}_i(n_1) \). Analogously to above, at \( \{\tilde{m}(n_1)|_{i-1}\} \), the expected utility from \( s_i \) has to be greater or equal than the expected utility from any \( s'_i \in S_i\{\tilde{m}(n_1)|_{i-1}\} \) such that \( s'_i\{\tilde{m}(n_1)|_{i-1}\} = m_i^+(\tilde{m}(n_1)|_{i-1}) \), implying that 
\[
\sum_{l \in \{0, n_i\}, \theta_{-1} \in \Theta_{-1}} (v(\theta) q^1_l + t^1_l) \lambda\{(l, \theta_{-1})\} > 0
\]
for some \( \lambda \in \Delta(\{0, \tilde{m}_i(n_1)\} \times \Theta_{-1}) \), which can hold only if \( \tau_i(\tilde{m}_i(n_1)) = \theta_i \).

b) It suffices to show that for any \( m_i \in \{1, \ldots, n_i\} \) such that \( (m_{i-1}, m_i) \in H^{-i} \), \((s_i, \theta_i) \in F_i^\infty \) and \( s_i\{m_{i-1}\} = m_i \) imply \( \tau_i(m_i) = \theta_i \). But this is immediate from lemma 5 b), as \( \mu_i(S_{-i} \times \{\tau_1(m_1), \ldots, \tau_{i-1}(m_{i-1})\}|\{m_{i-1}\}) = 1 \) for all \( \mu_i \in \Phi_i^\infty \), so that \( i \)'s expected valuation must be in \( V_i(\theta_i||m_{i-1}) \).

c) We show that \( F^\infty \subseteq \mathcal{F}(m_1, \ldots, 1), m_i \in H^{=i} \) and \( m_i \neq m_i^+(m_{i-1}) \) imply \( F^\infty \subseteq \mathcal{F}(m_{i-1}, m_i^+(m_i), n_{i+1} + 1, \ldots, n_i + 1) \). The claim then follows by an inductive argument analogous to the one made in the proof of a).

Suppose that \((s_i, \theta_i) \in \rho_i(\mu_i), \mu_i \in \Phi_i^\infty \), is such that \( s_i\{m_{i-1}\} = m_i^+(m_i) \), but that \( \tau_i(m_i^+(m_i)) \neq \theta_i \). If \( \mu_i(S_{-i} \times \{\theta_{-i} \in \Theta_{-i}; \nu_i(\theta) \geq v^i m^+(m_i)|\{m_{i-1}\}\})|\{m_{i-1}\} = 0 \) then \( m_i^+(m_i) \neq m_i(m_{i-1}) \) and playing \( s_i \) leads to a (weakly) negative expected utility for \( \theta_i \) at \( \{m_{i-1}\} \) by lemma 5 a). Since any strategy prescribing \( m_i^+(m_i) \) yields strictly positive expected utility, a contradiction to the sequential rationality of \( s_i \) obtains. If \( \mu_i(S_{-i} \times \{\theta_{-i} \in \Theta_{-i}; \nu_i(\theta) \geq v^i m^+(m_i)|\{m_{i-1}\}\})|\{m_{i-1}\} > 0 \), let \( \tilde{m}_i \) be the smallest \( m_i \in \{1, \ldots, n_i\} \) for which \( v^i m^i \) is in \( V_i(\theta_i||m_{i-1}) \) and \( v^i m^i > v^i m^i(m_i) \), and let \( s'_i \in S_i\{m_{i-1}\} \) be such that \( s'_i\{m_{i-1}\} = \tilde{m}_i \). Using the supposition \( F^\infty \subseteq \mathcal{F}(m_1, \ldots, 1) \),
\[
U^\mu_i(s'_i, \theta_i, \{m_{i-1}\}) = \sum_{l \in \{\tilde{m}_i, \ldots, n_i\}} (v^i q^i_l + t^i_l) \mu_i(S_{-i} \times \{\theta_{-i} \in \Theta_{-i}; \nu_i(\theta) = v^i_l\}|\{m_{i-1}\})
\]
If \( m_i^+(m_i) = m_i^+(m_{i-1}) \), this is strictly greater than \( U^\mu_i(s_i, \theta_i, \{m_{i-1}\}) \) (note that in this case \( V_i(\theta_i||m_{i-1}) \subseteq \{v^i m^i, \ldots, v^i m^i\} \) and thus \( s'_i \) yields the “best possible” expected utility, while by lemma 5 b) and \( \tau_i(m_i^+(m_{i-1})) \neq \theta_i \) the expected utility from \( s_i \) has to be strictly smaller).
\[ m_i^\dagger(m_i) \neq m_i(m_{i-1}) \text{ there exists a } \lambda \in \Delta(\{0, \ldots, n_i\} \times \Theta_{-i}) \text{ with the same marginal on } \Theta_{-i} \text{ as } \mu_i \text{ such that} \]
\[
U_i^\mu(s_i, \theta_i, \{m_{i-1}\}) = \sum_{l \in \{0\} \cup \{m_i^\dagger(m_i), \ldots, n_i\} \setminus \{l; \nu_i^l \in V_i(\theta_i||m_{i-1})\}, \theta_{-i} \in \Theta_{-i}; v_{i} \geq v_{i}^\mu_i} (v_i(\theta)q_i^l + t_i^l)\lambda\{(l, \theta_{-i})\}
+ \sum_{l \in \{0\} \cup \{m_i^\dagger(m_i), \ldots, n_i\} \setminus \{l; \nu_i^l \in V_i(\theta_i||m_{i-1})\}, \theta_{-i} \in \Theta_{-i}; v_{i}^\mu_i > v_{i}^\nu_i} (v_i(\theta)q_i^l + t_i^l)\lambda\{(l, \theta_{-i})\}.
\]

The first sum is strictly smaller than \( U_i^\mu(s_i', \theta_i, \{m_{i-1}\}) \) by lemma 5 b) and the second sum is (weakly) negative by lemma 5 a) because \( \{\theta_{-i} \in \Theta_{-i}; v_{i}^\mu_i > v_{i}^\nu_i\} = \emptyset \). Contradiction to \( s_i \) being sequentially rational. \( \square \)

From lemma 7 we can conclude that \( F^\infty \subseteq F(1, \ldots, 1) = F \). Since \( V_i(\theta_i||m_{i-1}) \subseteq V_i(\theta_i) \) for all \( m_{i-1} \in H_i \) and by assumption \( V_i(\theta_i) \cap V_i(\theta_i') = \emptyset \) for all \( \theta_i, \theta_i' \in \Theta_i, \theta_i \neq \theta_i' \), this means that \( \Gamma \) strategically distinguishes all payoff type profiles. This completes the proof of proposition 3. \( \square \)

We can view a valuation function \( v_i : \Theta \rightarrow \mathbb{R} \) as a point in \( \mathbb{R}^{\#\Theta} \), and a profile of valuation functions \( v = (v_1, \ldots, v_I) \) as a point in \( \mathbb{R}^{I\#\Theta} \). Then the set
\[
\mathcal{Y} = \{v \in \mathbb{R}^{I\#\Theta}; \forall i \in I, \theta_i, \theta_i' \in \Theta_i, \theta_i \neq \theta_i' : V_i(\theta_i) \cap V_i(\theta_i') = \emptyset\}
\]
is open and its complement has Lebesgue measure zero\(^\dagger\). Hence we can call \( \mathcal{Y} \) generic, and propositions 2 and 3 imply the following corollary.

**Corollary 1** In private consumption environments with generic valuation functions, social choice function \( f \) is rv-implementable if and only if \( f \) is epIC.

It is clear from the proof of proposition 3 that some payoff types of some agents can share an ex-post valuation but still be strategically distinguishable. The sufficient conditions of proposition 3 can be weakened to the requirement that

- \( \exists i \in I \) such that \( V_i(\theta_i) \cap V_i(\theta_i') = \emptyset \) for all \( \theta_i, \theta_i' \in \Theta_i, \theta_i \neq \theta_i' \), and
- \( \forall m_i \in \{1, \ldots, n_i\}; \exists j = j(m_i), j \neq i \), such that \( V_j(\theta_j||m_i) \cap V_j(\theta_j'||m_i) = \emptyset \) for all \( \theta_j, \theta_j' \in \Theta_j, \theta_j \neq \theta_j' \), and
- \( \forall m_i \in \{1, \ldots, n_i\}; \forall m_j \in \{1, \ldots, n_j(m_i)\} \), such that \( V_{j(m_i)}^m(\theta_j||m_i) \) for some \( \theta_j \in \Theta_{j(m_i)} \), \( \exists k = k(m_i, m_j), k \neq i, j(m_i) \), such that \( V_k(\theta_k||(m_i, m_j)) \cap V_k(\theta_k'||(m_i, m_j)) = \emptyset \) for all \( \theta_k, \theta_k' \in \Theta_k, \theta_k \neq \theta_k' \), and

\(^\dagger\)The set is a subset of \( \{v \in \mathbb{R}^{I\#\Theta}; v_i(\theta) \neq v_j(\theta') \) for all \( i, j \in I, \theta, \theta' \in \Theta\}. \)

28
etc.

where $V_j(\theta_j|m_i)$ is the set of $\theta_j$’s valuations that can arise from point beliefs when $j$ knows the payoff type of $i$ to be $\tau_i(m_i)$, and $V_k(\theta_k|(m_i, m_j))$ etc. are defined similarly. The next example demonstrates that even this relaxed sufficient conditions are not necessary for strategic distinguishability. In the example, for every agent the sets of ex-post valuations of all of the agent’s payoff types intersect. However, all payoff type profiles are strategically distinguishable.

**Example 5.1** Let $I \geq 3$, $\Theta_i = \{0, 1\}$ and $v_i(\theta) = \theta_i + \frac{1}{I-1} \sum_{j \neq i} \theta_j$ for all $i \in I$. Then for any $i \in I$, the set of ex-post valuations is

$$V_i(0) = \{0, \frac{1}{I-1}, \frac{2}{I-1}, \ldots, 1\}$$

for payoff type 0 and

$$V_i(1) = \{1, \frac{I}{I-1}, \frac{I+1}{I-1}, \ldots, 2\}$$

for payoff type 1. Since $V_i(0) \cap V_i(1) = \{1\}$, even the weakened sufficient conditions for strategic distinguishability of all payoff type profiles are violated. The mechanism $\Gamma$ defined in appendix C and depicted for the case $I = 3$ in figure 6\(^\dagger\) nonetheless strategically distinguishes all payoff type profiles: First, truth-telling at every node is strongly rationalizable for every payoff type of every agent. (Truth-telling at every node is a best response to the belief that others told and/or will tell the truth at every node, as well. This is true independent of what the marginal on others’ payoff types is at any node.) Second, truth-telling at every node is the unique strongly rationalizable strategy of for every payoff type of every agent, as we can see as follows:

\[\text{Figure 6: } \Gamma \text{ that strategically distinguishes all payoff type profiles}\]

\[\text{\footnotesize \begin{tabular}{c c c c c}
 0 & 1 & 2 & 0 & 2 \\
 1 & 3 & 0 & 1 & 3 \\
 0 & 1 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 & 0 \\
 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\
 1.5 & 1.5 & 1.5 & 1.5 & 1.5
\end{tabular}}\]

\(^\dagger\)The $i$-th line beneath the mechanism lists the options allocated to agent $i$. The options $o^n$ are defined as in the proof of proposition 3, but we omit the agent subscripts here.
such that $f^\bar{s}$ satisfies all payoff type profiles.

**Proof**. A.1 Proof of Lemma 2

The final proposition shows that even dynamic mechanisms cannot always strategically distinguish all payoff type profiles.

**Proposition 4** For any $i \in I$, $\theta_i, \theta_i' \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$, if $v_i(\theta_i, \theta_{-i}) = v_i(\theta_i', \theta_{-i})$ then $(\theta_i, \theta_{-i}) \sim (\theta_i', \theta_{-i})$.

**Proof.** Take any mechanism $\Gamma$ and pick an arbitrary $s_{-i} \in R^\infty_{-i}(\theta_{-i})$. Let $\delta_{(s_{-i}, \theta_{-i})} \in \Delta(\Sigma_{-i})$ denote the point belief in $(s_{-i}, \theta_{-i})$ and let $\mu_i' \in \Phi^\infty_i$. Define $\mu_i : 2^{\Sigma_i} \times \mathcal{H}_i \to [0, 1]$ by $\mu_i(\cdot|\mathcal{H}) = \delta_{(s_{-i}, \theta_{-i})}$ for $\mathcal{H} \in \mathcal{H}_i(s_{-i})$ and $\mu_i(\cdot|\mathcal{H}) = \mu_i'(|\mathcal{H})$ for $\mathcal{H} \notin \mathcal{H}_i(s_{-i})$. Note that $\mu_i$ is a CPS, and, since $\mu_i' \in \Phi^\infty_i$ and all the mass of $\delta_{(s_{-i}, \theta_{-i})}$ concentrates on strongly rationalizable strategy-payoff type pairs, $\mu_i \in \Phi^\infty_i$. Let $s_i' \in r_i(\theta_i', \mu_i)$, then there exists $s_i \in r_i(\theta_i, \mu_i)$ such that $s_i|_{\mathcal{H}_i(\theta_{-i})} = s_i'|_{\mathcal{H}_i(s_{-i})}$. In summary, $(s_{-i}, \theta_{-i}) \in F^\infty_{-i}$, $(s_i, \theta_i), (s_i', \theta_i') \in F^\infty_i$ and $\zeta(s_i, s_{-i}) = \zeta(s_i', s_{-i})$. \qed

## A Proofs

### A.1 Proof of Lemma 2

**Proof.** Since $\Sigma_{-i}$ and $\mathcal{H}_i$ are finite we can view $\Delta^\mathcal{H}_i(\Sigma_{-i})$ as a subspace of the $\#\Sigma_{-i} \cdot \#\mathcal{H}_i$-dimensional Euclidean space. For any $i \in I$, $k \in \mathbb{N}$, $(s_i, \theta_i) \notin F_i^{k+1}$ and $\mu_i \in \Phi^k_i$ there exist $\mathcal{H} \in \mathcal{H}_i(s_i)$ and $s_i' \in S_i(\mathcal{H})$ such that $U^\mu_i(s_i', \theta_i, \mathcal{H}) = U^\mu_i(s_i, \theta_i, \mathcal{H}) > 0$. By continuity of $f^{(i,k,s_i,\theta_i,\mu_i)} : \Delta^\mathcal{H}_i(\Sigma_{-i}) \to \mathbb{R}, \mu_i' \mapsto U^\mu_i(s_i', \theta_i, \mathcal{H}) - U^\mu_i(s_i, \theta_i, \mathcal{H})$ there is $\varepsilon(i, k, s_i, \theta_i, \mu_i) > 0$ such that $f^{(i,k,s_i,\theta_i,\mu_i)}$ assumes a strictly positive minimum $\eta_i(i, k, s_i, \theta_i, \mu_i)$ on the closed ball $B_{\varepsilon(i, k, s_i, \theta_i, \mu_i)}(\mu_i)$ with radius $\varepsilon(i, k, s_i, \theta_i, \mu_i)$ and center $\mu_i$. Since $\Phi^k_i$ is compact,
the cover \((B_{\varepsilon(i,k,s_i,\theta_i,\mu_i)}(\mu_i))_{\mu_i \in \Phi_i}\) of open balls \(B_{\varepsilon(i,k,s_i,\theta_i,\mu_i)}(\mu_i)\) has a finite subcover \((B_{\varepsilon(i,k,s_i,\theta_i,\mu_i^m)}(\mu_i^m))_{m=1}^{n(i,k,s_i,\theta_i)}\). As there are only finitely many distinct \(F^{k+1}\), it now suffices to let

\[
\eta_T = \min_{i \in I, k \in \mathbb{N}} \min_{(s_i, \theta_i) \notin F^{k+1}} \eta_T(i, k, s_i, \theta_i, \mu_i^m).
\]

\(\square\)

A.2 Proof of Lemma 3

Proof. By induction on \(k\). Obviously a) and b) are true for \(k = 0\). Now take \(k \geq 1\) and suppose a) and b) are true for all \(k' < k\). Together with \(\mathcal{H}_i^*(F_{\ast}^{s,\infty}) = \mathcal{H}_i^*\) for all \(i \in I\), a) for \(k' < k\) implies

\[
F_{-i}^{k,k'} \cap \Sigma_{-i}(\mathcal{H}_i^*) \neq \emptyset \iff F_{-i}^{k'} \cap \Sigma_{-i}(\{\theta_i, \mathcal{H}_i^*\}) \neq \emptyset, \quad \forall i \in I, k' < k, \theta_i \in \Theta_i^L, \mathcal{H}_i^* \in \mathcal{H}_i^*, (4)
\]

so that the highest degree \(k'\) of rationality that \(i\) can ascribe to \(-i\) coincides at \([\theta_i, \mathcal{H}_i^*]\) and \(\mathcal{H}_i^*\).

a) For any \(i \in I\) and any strategy \(s_i^* \in R_i^{s,\infty}(\theta_i)\) there is a CPS \(\mu_i^* \in \Phi_i^{s,\infty}\) to which \(s_i^*\) is sequential best response for payoff type \(\theta_i\), \(s_i^*, \theta_i \in \rho_i^l(\mu_i^*)\). Define \(\mu_i : 2^{\Sigma_i-1} \times \mathcal{H}_i \rightarrow [0,1]\) such that \(\mu_i((s_{-i}, \theta_{-i})|\{\emptyset\}) = \mu_i((s_{-i}, \theta_{-i})|\mathcal{H}_i^0) = \mu_i^*((\phi_{-i}^\theta)^{-1}(s_{-i}) \times \{\theta_{-i}\}|\emptyset)\) and

\[
\mu_i((s_{-i}, \theta_{-i})|\{\theta_i, \mathcal{H}_i^*\}) = \mu_i^*((\phi_{-i}^\theta)^{-1}(s_{-i}) \times \{\theta_{-i}\}|\mathcal{H}_i^*), \quad \forall (s_{-i}, \theta_{-i}) \in \Sigma_{-i}, \theta_i \in \Theta_i^L, \mathcal{H}_i^* \in \mathcal{H}_i^*.
\]

\(\mu_i\) is a CPS and by (4) and a) for \(k' < k\) an element of \(\Phi_{i}^{k+1}\). At each \(\mathcal{H} \in \mathcal{H}_i, U_i^{\mu_i}(\cdot, \theta_i, \mathcal{H})\) is the sum of four terms, each of which \(\phi_i^0(s_i^*)\) maximizes in \(S_i(\mathcal{H})\). First, the expected utility from \((1 - \delta - \delta^2)\sum_{i=1}^{L} f((s_{i}(\mathcal{H}_i^0)^1))_{i \in I}\), which \(\phi_i^0(s_i^*)\) maximizes because \(f\) is epiC.† Second, the expected utility from \(\delta C^\ast(j\ast(\varphi_1(s_1), \ldots, \varphi_l(s_l)))\), which \(\phi_i^0(s_i^*)\) maximizes because \(s_i^*\) is a sequential best response for \(\theta_i\) with respect to \(\mu_i^*\). Third, the expected utility from \(\delta^{2\frac{1}{L}}\sum_{j=1}^{L} r_j(\zeta(s))\), which \(\phi_i^0(s_i^*)\) maximizes because \(r_i(\zeta(\phi_i^0(s_i^*)|\cdot)) = z_i\): if agent \(i\) plays \(\phi_i^0(s_i^*)\) he submits \(L\) times his true payoff type \(\theta_i, \theta_i\), and plays \(s_i^* \in R_i^{s,\infty}(\theta_i)\) in the “\(\Gamma^\ast\)-part”. No \(h^* \notin H^\ast(R_i^{s,\infty}(\phi_i^0(s_i^*)|\mathcal{H}_i^0))\) can result in the “\(\Gamma^\ast\)-part”, no matter what the other agents play, and there is no \(m\) such that \(\phi_i^0(s_i^*)|\mathcal{H}_i^0)^m \neq \phi_i^0(s_i^*)|\mathcal{H}_i^0)^{1}\). Fourth, the expected utility from \(\delta^{2\frac{1}{L}}\sum_{j=1}^{L} r_j(\zeta(s))\), which \(\phi_i^0(s_i^*)\) maximizes because \(r_j(\zeta(s)) = z_j\) for any \(s_i \in S_i\) and any strategy profile \(s_{-i}\) that \(i\) expects to be played with strictly positive probability. Therefore \((\phi_i^0(s_i^*), \theta_i) \in \rho_i(\mu_i)\).

†We say that \(\phi_i^0(s_i^*)\) maximizes the expected utility from \(g(s)\) in \(S_i(\mathcal{H})\) (for payoff type \(\theta_i\) with beliefs \(\mu_i\), where \(g : S \rightarrow \mathbb{R}^{s,X}\), if

\[
\phi_i^0(s_i^*) \in \arg\max_{s_i \in \mathcal{S}_i(\mathcal{H})} \sum_{(s_{-i}, \theta_{-i}) \in \Sigma_{-i}(\mathcal{H})} u_i(g(s), \theta)\mu_i((s_{-i}, \theta_{-i})|\mathcal{H}).
\]
b) For any $i \in I$ and any $s_i \in R^k_i(\theta_i)$ there is $\mu_i \in \Phi_{i}^{k-1}$ such that $(s_i, \theta_i) \in \rho_i(\mu_i)$. Define $\mu^*_i : 2^{\Sigma_{-i}} \times \mathcal{H}^*_i \to [0, 1]$ such that

$$\mu^*_i((s^*_i, \theta_{-i})|\mathcal{H}^*) = \mu_i(\varphi^{-1}_i(s^*_i) \times \{\theta_{-i}\}|[s_i(\mathcal{H}^0_i), \mathcal{H}^*]), \quad \forall(s^*_i, \theta_{-i}) \in \Sigma_{-i}, \mathcal{H}^* \in \mathcal{H}^*_i$$

(where $[s_i(\mathcal{H}^0_i), \{\emptyset\}]$ designates $\mathcal{H}^0_i$ if $\emptyset \notin \mathcal{H}^*_i$). Then $\mu^*_i$ is a CPS and by (4) and b) for $k' < k$ an element of $\Phi_{i}^{k,k-1}$. Suppose now that $(\varphi_i(s_i), \theta_i) \notin F_{i}^{*,k}$ then by lemma 2 there are $\mathcal{H}^* \in \mathcal{H}^*_i(\varphi_i(s_i))$ and $s^*_i \in S^*_i(\mathcal{H}^*)$ such that $U_{i}^{\mu_i^*}(s^*_i, \theta_i, \mathcal{H}^*) > U_{i}^{\mu_i}(\varphi_i(s_i), \theta_i, \mathcal{H}^*) + \eta_{i}^*$. But then by (2), $s'_i \in S_i([s_i(\mathcal{H}^0_i), \mathcal{H}^*])$ defined by $s'_i(\mathcal{H}^0_i) = s_i(\mathcal{H}^0_i)$ and $s'_i(\theta_i, \mathcal{H}^*) = s^*_i(\mathcal{H}^*)$ for all $[\theta_i, \mathcal{H}^*_i]$ must give strictly higher expected utility under $\mu_i$ at $[s_i(\mathcal{H}^0_i), \mathcal{H}^*]$ than $s_i$; playing $s'_i$ instead of $s_i$ yields an expected utility gain of at least $\delta \eta_{i}^*$ in $\delta C^*(\zeta(s_i), ..., \varphi_i(s_i))$ and a possible expected utility loss of at most $\delta^2 C_0$ in $\delta^2 \frac{1}{k} \sum_{j \in I} r_j(\zeta(s))$. This contradicts $(s_i, \theta_i) \in \rho_i(\mu_i)$. Therefore $(\varphi_i(s_i), \theta_i) \in F_{i}^{*,k}$.  

\[ \square \]

### A.3 Proof of Proposition 3: Options and the Proof of Lemma 5

Let $(q^{m}, t^{m}) = \left(\frac{1}{T} - \frac{0.5}{T}(v^{m} + v^{-1}) \right)$, and let

$$q^m = \min_{l = m + 1, ..., n_i} \left\{ \frac{0.5\mu_i((q^l, t^l), \pi_i(v^l))}{v^l - 0.5(v^m + v^{m-1})} \right\},$$

$$t^m = -0.5q^m(v^m + v^{m-1}),$$

$m = n_i - 1, ..., 1$. Moreover, let $(q^0, t^0) = (0, 0)$ and $(q^m, t^m)$ for $m \in \{0, ..., n_i\}$. Then:

a) $0.5(v^m + v^{-1})q^m + t^m = 0$ for $m \in \{1, ..., n_i\}$

By definition of $t^m$ for $m \in \{1, ..., n_i\}$, and by definition of $t^m$ and $q^m$ for $m = n_i$.

b) $q^m \in (0, \frac{1}{T})$ for $m \in \{1, ..., n_i\}$

By definition, $q^m > 0$. For $m \in \{1, ..., n_i - 1\}$, if $q^m > 0$ for all $l \in \{m + 1, ..., n_i\}$, then both $v^l - 0.5(v^m + v^{m-1}) > 0$ and $u_i((q^l, t^l), \pi_i(v^l)) = v^l q^l + t^l > 0.5(v^m + v^{m-1})q^m + t^m = 0$ for all $l \in \{m + 1, ..., n_i\}$, and therefore $q^m > 0$. By definition, $q^m \leq \frac{1}{T}$. For $m \in \{1, ..., n_i - 1\}$,

$$q^m \leq \frac{0.5\mu_i((q^m, t^m), \pi_i(v^m))}{v^m - 0.5(v^m + v^{m-1})} < \frac{u_i((q^m, t^m), \pi_i(v^m))}{v^m - 0.5(v^m + v^{m-1})} = \frac{1}{T}.$$

$$q^m q^m + t^m > 0 \text{ for } m \in \{1, ..., n_i\}$$

---

1We let $u_i((q, t, \theta) = u_i((0, 0, q, 0, ..., 0), (0, ..., 0, t, 0, ..., 0), \theta)$ for all $(q, t) \in \mathbb{R}$. 

32
By b) and the definition of $q_i^0$, any profile of options $(o_i^{m_1}, \ldots, o_i^{m_I})$ with $m_i \in \{0, \ldots, n_i\}$ for all $i \in I$ is an element of $Y$ and thus can be assigned as the outcome of a mechanism (if $v_i^m \not\in [-B, B]$ for some $i \in I$ and some $m_i \in \{0, \ldots, n_i\}$, redefine $o_i^m$ as $\frac{1}{K} o_i^m$ for all $i \in I$, $m_i \in \{0, \ldots, n_i\}$ and some sufficiently large $K > 0$).

We can now prove lemma 5:

Proof. a) Let $l \in \{1, \ldots, m - 1\}$, then $v_l^i q_i^m + t_i^m < 0.5(v_i^m + v_i^{m-1}) q_i^m + t_i^m = 0$.

b) The claim is true for $l = 0$ because $v_l^m q_i^m + t_i^m > 0$. For $l \in \{m+1, \ldots, n_i\}$, $v_l^m q_i^m + t_i^m > 0 > v_i^m q_i^m + t_i^m$ by a). For $l \in \{1, \ldots, m - 1\}$, by the definition of $q_i^l$,

$$q_i^l \leq \frac{0.5 u_i((q_i^m, t_i^m), \pi_i(v_i^m))}{v_i^m - 0.5(v_l^i + v_l^{i-1})}$$

and therefore $q_i^l v_i^m - 0.5 q_i^l (v_l^i + v_l^{i-1}) = u_i(q_i^l, \pi_i(v_i^m)) < u_i(q_i^m, \pi_i(v_i^m))$. □

\section*{B The Surjection of Remark 1}

We are going to define a utility-preserving surjection $g$ from $\bar{Y}$ to $Y$.  \footnote{\textit{g} is not bijective. For example, if $I = 2$, then $g$ maps both $\bar{y}$ such that $\bar{y}(1, 0, -B) = \bar{y}(0, 1, -B, B) = 0.5$ and $\bar{y}'$ such that $\bar{y}'(1, 0, -B, B) = \bar{y}'(0, 1, B, -B) = 0.5$ to $(q, t) = (0.5, 0.5, 0, 0)$. This does not matter, however, as for any $\bar{y} \in \bar{Y}$, all $i \in I$ are indifferent between any member of $g^{-1}(\bar{y})$.} Let $g : \bar{Y} \to Y$ map $\bar{y}$ to the $(q, t)$ for which

$$q_i = \sum_{(\bar{q}, \bar{t}) \in X} \bar{y}(\bar{q}, \bar{t}) 1_{\{\bar{q}_i = 1\}}(\bar{q}) \quad \text{and} \quad t_i = \sum_{(\bar{q}, \bar{t}) \in X} \bar{y}(\bar{q}, \bar{t}) \bar{t}_i, \quad \forall i \in I,$$

where $1_A, A \subseteq \bar{Y}$, is the indicator function of $A$. Then for every $i \in I$ and every $\bar{y} \in \bar{Y}$,

$$\bar{u}_i(\bar{y}, \theta) = v_i(\theta) \sum_{(\bar{q}, \bar{t}) \in X} \bar{y}(\bar{q}, \bar{t}) \bar{q}_i + \sum_{(\bar{q}, \bar{t}) \in X} \bar{y}(\bar{q}, \bar{t}) \bar{t}_i = u_i(g(\bar{y}), \theta).$$

Moreover, let $(q, t) \in Y$. Let $q_0 = 1 - \sum_{i \in I} q_i$, and for each $i \in I$ let $r_i \in [0, 1]$ solve $t_i = (-B)r_i + B(1 - r_i)$. Define $\bar{y} \in \bar{Y}$ by

$$\bar{y}(\bar{q}, \bar{t}) = \left(1_{\{\bar{q}=(0,\ldots,0)\}}(\bar{q})q_0 + \sum_{i \in I} 1_{\{\bar{q}_i = 1\}}(\bar{q})q_i\right) \prod_{i \in I} \left(1_{\{\bar{t}_i = -B\}}(\bar{q})r_i + 1_{\{\bar{t}_i = B\}}(\bar{q})(1 - r_i)\right).$$
Then \( g(\bar{y}) = (q', t') \) for
\[
q'_i = \sum_{i \in \{B, -B\}^I} q_i \prod_{i \in I} (1_{\{i, = -B\}} r_i + 1_{\{i, = B\}} (1 - r_i)) = q_i;
\]
\[
t'_i = \sum_{i \in \{B, -B\}^I} \prod_{i \in I} (1_{\{i, = -B\}} r_i + 1_{\{i, = B\}} (1 - r_i)) \tilde{t}_i
\]
\[= (-B)r_i + B(1 - r_i) = t_i.
\]
That is, \( g(\bar{y}) = (q, t) \).

C  Mechanism of Example 5.1

Define \( \Gamma = \langle H, (\mathcal{H}_i)_{i \in I}, P, C \rangle \) as follows. The agents publicly and in sequence announce their payoff types:
\[
H = \{ h \in \mathcal{F}; h \preceq h' \text{ for some } h' \in \{0, 1\}^I \},
\]
where \( \mathcal{F} \) is the set of finite sequences with codomain \( \{0, 1\} \), the player function is \( P : H \setminus T \to I \) such that \( P(h) = i \) if \( l_i = i - 1 \), for all \( i \in I \) and \( h \in H \), and \( \mathcal{H}_i = \{\{h\}; h \in H_i\} \). The outcome function \( C : T \to Y \) maps \( h \) to the lottery \( C(h) = (o_1, \ldots, o_I) \) such that
\[
o_1 = \begin{cases}
\sigma_{1^{-1}}^{I} & \text{if } h_1 = 1 \text{ and } (h_2, \ldots, h_I) \neq (0, \ldots, 0), \\
\sigma_{1} & \text{if } h = (1, 0, \ldots, 0), \\
\sigma_{0} & \text{if } h_1 = 0
\end{cases}
\]
\[
o_i = \begin{cases}
\sigma_{1^{-1}}^{I} & \text{if } h_1 = 1 \text{ and } h_i = 1 \\
\sigma_{1} & \text{if } (h_1 = 1 \text{ and } h_i = 0) \text{ or } (h_1 = 0 \text{ and } h_i = 1), \\
\sigma_{2^{-1}}^{I} & \text{if } h_1 = 0 \text{ and } h_i = 0
\end{cases}
\]
and
\[
o_I = \begin{cases}
\sigma_{1^{-1}}^{I} & \text{if } h_1 = 1, h_I = 1 \text{ and } (h_2, \ldots, h_{I-1}) \neq (0, \ldots, 0) \\
\sigma_{1} & \text{if } (h_1 = 1, h_I = 0 \text{ and } (h_2, \ldots, h_{I-1}) \neq (0, \ldots, 0)) \text{ or } \\
h = (1, 0, \ldots, 0, 1) \text{ or } (h_1 = 0 \text{ and } h_I = 1) \\
\sigma_{2^{-1}}^{I} & \text{if } h_1 = 0 \text{ and } h_I = 0 \\
\sigma_{1^{-1}}^{I} & \text{if } h = (1, 0, \ldots, 0, 0)
\end{cases}
\]
where \( i \in \{2, \ldots, I - 1\} \) and \( \sigma^m \) is defined as in the proof of proposition 3 (we omit subscripts since the options are the same for all agents).
References


